World Wide Web (2013) 16:31–72
DOI 10.1007/s11280-011-0155-z

Social networks on the Internet

Katarzyna Musiał · Przemysław Kazienko

Received: 12 November 2009 / Revised: 28 June 2011
Accepted: 27 December 2011 / Published online: 26 January 2012

Abstract The rapid development and expansion of the Internet and the social–based services comprised by the common Web 2.0 idea provokes the creation of the new area of research interests, i.e. social networks on the Internet called also virtual or online communities. Social networks can be either maintained and presented by social networking sites like *MySpace*, *LinkedIn* or indirectly extracted from the data about user interaction, activities or achievements such as emails, chats, blogs, homepages connected by hyperlinks, commented photos in multimedia sharing system, etc. A social network is the set of human beings or rather their digital representations that refer to the registered users who are linked by relationships extracted from the data about their activities, common communication or direct links gathered in the internet–based systems. Both digital representations named in the paper internet identities as well as their relationships can be characterized in many different ways. Such diversity yields for building a comprehensive and coherent view onto the concept of internet–based social networks. This survey provides in–depth analysis and classification of social networks existing on the Internet together with studies on selected examples of different virtual communities.

Keywords online social network · taxonomy of social networks · internet identity · virtual identity · internet relationship · homogeneous social network · system-based social network · internet multisystem social network · social networking site · survey

K. Musiał (✉)
School of Natural and Mathematical Sciences, King's College London, Strand Campus, London WC2R 2LS, UK
e-mail: katarzyna.musial@kcl.ac.uk

P. Kazienko
Institute of Informatics, Wrocław University of Technology, Wybrzeze Wyspianskiego 27, 50-370 Wroclaw, Poland
e-mail: przemyslaw.kazienko@pwr.wroc.pl

1 Introduction

The concept of social network, first coined in 1954 by J. A. Barnes in [4], has been in a field of study of modern sociology, anthropology, geography, social psychology, organizational studies and computer science for last few decades.

The person who created the modern social network paradigm was Stanley Milgram. He studied the small-world problem that can be described by means of indirect relationships. Even if two persons x and y do not know each other directly, they can share a mutual relationship that is another person who knows them both [52, 67]. The theoretical model of this small-world problem was created by Pool and Kochen [52] and served as the basic for Milgram's research that was purely pictorial. Stanley Milgram conducted two experiments – Kansas Study and Nebraska Study – in which he asked many people from one city to forward a letter to a chosen person in another remote city. It should have been done only by giving this letter to a person that these people knew on a first–name basis. Afterward, he analyzed forwarding paths of the letters and concluded that people in the USA form a social network and they are connected to this network with "six degrees of separation". It means that a message in such a network would be delivered by average five intermediaries [52]. Kochen confirmed that this value is relatively even if the starter selection criteria are changed [16]. Howard claims that six degrees of separations may be true offline while less than three degrees is more likely online [31].

Since 1967 social networks have become one of the research areas where scientists from different fields are looking for inspiration. Thus, social networks and especially social network analysis (SNA) supported by computer science provide the opportunity to expand other branches of knowledge.

The concept of social network and social network analysis have been developed for many domains such as corporate partnership networks (law partnership) [47], scientist or other professions collaboration networks [19, 58], family networks, friendship network of students [2], company director networks [64], sexual contact networks [56], customer networks [24, 75], labour market [55], public health [11], psychology [60], etc. Recently, it becomes a part of the new discipline of science called *computational social science* [48].

With the expansion of the Internet and the increasing popularity of social and collaborative computing, recently commonly called social computing [61, 70], social networks have emerged as a significant and promising field of study within computer science. Social computing involves such activities as collecting, extracting, accessing, processing, computing, and visualizing of all kind of social information [40].

This survey has been created in order to define and classify social networks on the Internet, which have always existed in this worldwide communication medium. Nowadays, social networks can easily be extracted, often even from publicly available data. Based on this data, we can discover new knowledge about humans using internet services and the communities they are active in. Additionally, the profiles of social networks in the virtual world differ from the profiles of social networks existing in the real world and because of that they yield for more precise investigation including survey studies.

In this article, authors focus on describing and defining social networks that can be extracted from data available on the Internet as well as their different components: nodes and connections between them. Section 2 briefly introduces the concept of social networks that exist in the real world whereas in Section 3, a systematic approach was proposed to

distinguish different types of social networks on the Internet starting from the simplest one – homogenous social network, then moving to system-based social network and ending with the most complex one – internet multisystem social network. In the third Section, also the related work connected with social networks on the Internet is presented. After that, the concepts of node called internet identity and connection called internet relationship are presented in Section 4 and 5, respectively. The aim of Section 6 is to present possible classification of social networks on the Internet. The different examples of these networks are described in the following Section 7. The last sections of the survey are devoted to summarise the discussions and conclude the work.

2 What is a social network?

Before the social networks on the Internet will be defined, some basic ideas related to the social networks in general ought to be explained.

The general concept of society can be considered as the background for the social network definition. A society is not merely a simple aggregation of individuals; it is rather the sum of the relationships that connect these individuals to one another [51]. Hence, historically a social network is the finite set of nodes (actors) and edges (ties) that link these nodes, Figure 1.

Although the concept of social network appears to be quite obvious, almost every researcher describes it in a slightly different way. Some scientists define a social network in a very formal way, e.g. Yang, Dia, Cheng, and Lin [75] while others prefer more sociological approach [72]. More insight into the problem of the social network definition is presented in Table 1. The listing shows how the concepts of actor, relation, and social network are addressed by different researchers. Apparently, definitions from Table 1 are the most representative, although there are many other scientists who have investigated the general concept of social network [9, 16]. However, the other definitions are in fact a mixture of the presented ones.

3 Social networks on the internet

The continuously increasing popularity of the World Wide Web and the Internet caused that increasing number of types of services is available through a computer network. People who use these services have created a new kind of virtual societies usually called *online social networks*, e.g. [13, 31, 43, 49]. They can also be named as *web-based social networks* [24], *computer-supported social networks* [73] or *virtual communities* [10].

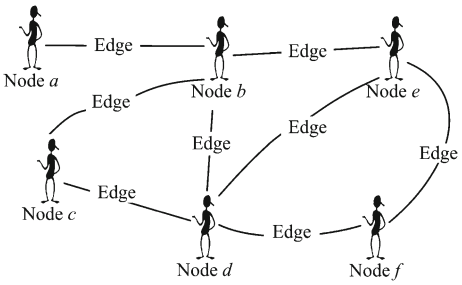

Figure 1 A social network consists of nodes and edges.

Table 1 Definitions of an actor, relation, social network and the examples of social networks.

Author	Actor	Relation	Social Network	Examples of SN
Wasserman and Faust [72]	An *actor* is a discrete individual, corporate or collective social units	A set of *ties* of a specific type; a tie is a *linkage* between a pair of actors	The finite set or sets of actors and one or more relations defined on them	Friendship among children in a classroom; all nations in the world and the formal diplomatic connections between them
Hanneman and Riddle [42]	*Actors* are also called *points*, *nodes* or *agents*	*Relationships*, *edges* or *ties*; one or more kinds of relations between pairs of actors	A set of actors that may have relationships with one another	Family; co-workers in a company; the network of neighbors; friendship among students in a classroom
Garton, Haythorntwaite, and Wellman [33]	*People*, *organizations* or other *social entities*	*Relationships*, such as *friendship*, *co-working* or *information exchange*	A set of social entities connected by a set of social relationships	Friendship among people; co-workers in a company; people who communicate with one another via computer
Hatala [44]	*Actors* are *people* or *groups of people*	*Patterns of interaction* or *ties* between actors	A set of actors with some patterns of interaction or "ties" between them; represented by graphs or diagrams illustrating the dynamics of the various connections and relationships within the group	Co-workers within a company
Liben-Nowell and Kleinberg [76]	*People* or *other entities* embedded in the social context	*Edges* represent *interaction*, *collaboration*, or *influence* between entities	*Structures* whose nodes represent entities embedded in the social context, and whose edges represent interaction, collaboration, or influence between entities	Co-authors of the scientific papers in a particular discipline; project groups in a large company; business leaders who have served together on a corporate board of directors
Yang, Dia, Cheng, and Lin [75]	A *node* in a graph; each node represents a *customer*	The undirected, unweighted *edges* in the graph; each edge represents the *connectedness* between two nodes	An undirected, unweighted graph	Customer's social network which is derived from customer's interaction data

The main features that distinguish social networks on the Internet from the social networks extracted based on interactions between people in the real world are as follows:

i. Lack of physical, in person contact – only by distance, sometimes very large distances.
ii. Usually the lack of unambiguous and reliable correlation between member's identity in the virtual community – internet identity (see Section 4) and their identity in the real world.
iii. The possibility of multimodal communication, simultaneously with many members; also the possibility of easy switches between different communication channels, especially online and offline, e.g. online VoIP and offline text communication.
iv. The simplicity of a break up and suspension of contacts or relationships.
v. The relatively high ease of gathering data about communication or common activities and its further processing.
vi. Potential lower reliability of the data about users available on the Internet. Users of internet services relatively frequently provide fake personal data due to privacy concerns.

The discussion about internet identity and human relationships – crucial components of internet-based social networks can be found in Sections 4 and 5, respectively, whereas the taxonomy of social networks on the Internet was placed in Section 6.

3.1 The concept of social networks on the internet

Since many different types of social networks can be distinguished on the Internet, let us try to specify some basic definitions of online social networks, beginning from the simplest homogeneous social network, through the system-based social network to the most complex Internet multisystem social network, Figures 2 and 3. More detailed insight into various social networks can be found in Section 7.

Definition 1 *A homogeneous social network on the Internet* $HSN=(IID,R)$ exists in the single internet-based system *S*. It consists of the finite set of internet identities *IID* – registered, non-anonymous users of system *S*, and the set of all social relationships *R* of the same kind that join pairs of *IID* members: $R=\{(iid_i,iid_j)\colon iid_i \in IID,\ iid_j \in IID\}$. *HSN* is also called *the single layer social network*.

The single system on the Internet is the homogeneous system maintained or operated by the same subject (company, group of companies, cooperating administrators) usually using

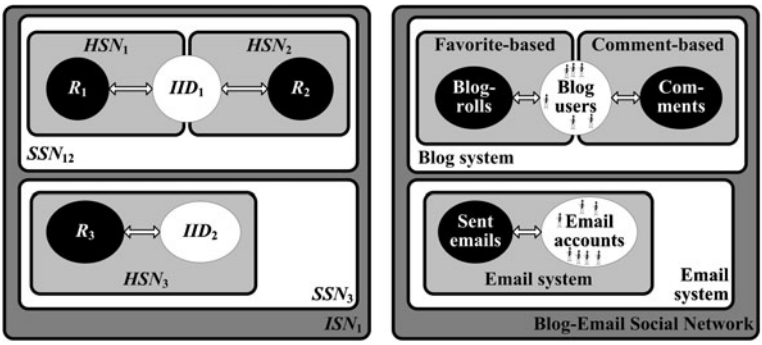

Figure 2 Homogeneous (*HSN*), system-based (*SSN*) and internet multisystem social network (*ISN*).

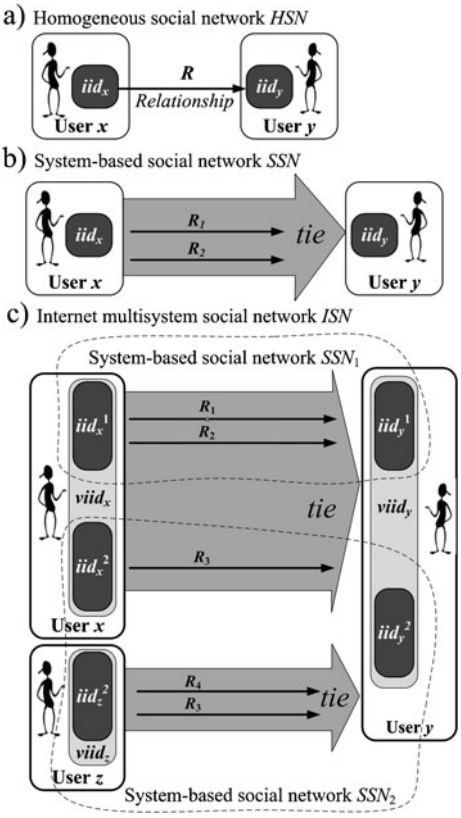

Figure 3 Homogeneous (*HSN*), system–based (*SSN*) and internet multisystem social network (*ISN*).

common interface and/or protocol. The main indicator of the system is the full availability of data. From this point of view, two email systems operated by two separate companies for example Microsoft (*Hotmail*) and Google (*Gmail*) are two different systems on the Internet unless they do not exchange internal data about user communication. Thus, we are unable to create any relationship between two *Gmail* users based on their email exchange, having the data only from the *Hotmail* server. In consequence, two separate *HSNs* have to be created, one from the data available for Microsoft and one based on Google's records. On the other hand, a social network built upon the personal homepages connected with one another with the help of HTML hyperlinks can be treated as a single system since the information about mutual relationships is public even though it is scattered.

An internet identity is a digital, authenticable and permanent representation of a person, organization or organizational unit, group of people, or other social entity like family or group of interest, for more details and examples see Section 4.

The same type of relationships means that two *HSN*'s members iid_i and iid_j share the same activity, e.g. they communicate with each other using emails or comment posts in the *WordPress* blogging system. Note that sometimes many different *HSN*s can be recognized within a single internet system. For example, based on the shared tagging lists to pictures, links to favourites and contact lists, three separate homogeneous social networks can be distinguished.

In general, relationships R can be either directed (as in definition 1) or undirected. In the latter case, the definition would have to be modified: $R=\{\{iid_i,iid_j\}: iid_i \in IID,\ iid_j \in IID\}$.

Besides, relationships can be either weighted ($R \rightarrow \Re$) or unweighted (binary) – all edges are considered equivalent. Due to social character of *HSN*, it is usually reasonable to ensure only irreflexive relationships, i.e. $(iid_i,iid_j) \in R \Rightarrow i \neq j$. In other words, self-choices relationships [72] are not considered. For more details about relationships on the Internet see Section 5.

Definition 2 *A system-based social network* $SSN=(IID,T)$ on the internet system S is a multilayered (multirelational) social network that consists of the finite set of internet identities *IID* and the set T of ties T linking pairs of internet identities. Set T is built from all relationships R_1, R_2, …, R_N existing in the system, where N is the number of homogeneous social networks *HSN* identified in the system.[1]

In other, more formal, words, set of ties is represented as: $T=\{(iid_i,iid_j,k_1,k_2,\ldots,k_N)$: $iid_i \in IID$, $iid_j \in IID$, $k_l=1 \Leftrightarrow (iid_i,iid_j) \in R_l$ or $k_l=0$ otherwise$\}$. The examples of internet systems, from which multilayered social networks can be extracted, are: blog systems (*WordPress*, *Blogger*), multimedia sharing systems (*Flickr*, *YouTube*), complex instant messengers (*Skype*, *ICQ*), *Gmail* email system extended with personalized searching by utilizing Google search engine. Each of these internet systems contains one or more *HSN*s, which form single *SSN*s.

In *WordPress*, users can both maintain their blogs as well as tag them with the keywords that usually provide the information about the content of their diaries. Based on tags used and shared by users, we can create a homogeneous social network HSN_1. All people who exploit the same tags get into mutual relationships R_1. On the other hand, *WordPress* users can also maintain a blogroll, i.e. a list of links to other blogs they like the most. These connections are the basis to create the second relationship type R_2 and another HSN_2 which can be called favourite–based homogeneous social network, Figure 2. Similarly to tags, the third R_3 and HSN_3 can be extracted from opinions that concern the same blog posts. People who comment the same blogs are in the mutual relationship and there is a high probability that users are interested in similar topics or prefer the same authors. Note that the relationships within different *HSN*s have different characteristics. More detailed analysis of a system–based social network has been performed for *Flickr* – a picture publishing system [35], see Section 7.5.

A homogenous social network is in fact simultaneously the system–based social network. Such case occurs when the set of ties T in the system–based social network is based only on one set of homogenous relationships R_1, e.g. homogeneous network HSN_3 extracted from email communication is, at the same time, the system–based (email–based) social network SSN_3, see Figure 2. However, it is valid only if we do not respect relationships derived from contacts in address books. In yet another example, personal web pages connected with hyperlinks form both the homogeneous and the system–based social network.

Some separate system-based social networks can be merged in one complex internet multisystem social network.

Definition 3 *An internet multisystem social network ISN* for the set of m system–based social networks $SSN_i=(IID_i,T_i)$, $i=1, \ldots, m$ is the tuple $(VIID^M,T^M)$, where $VIID^M$ is the set of virtual internet identities related to the same physical social entity.[2] T^M is the set of ties $T^M=\{(viid_x,viid_y,k_{11},\ldots, k_{1N_1},k_{21,\ldots,}k_{2N_2},\ldots,k_{m1},\ldots,k_{mN_m})$: $viid_x \in VIID^M$, $viid_y \in VIID^M$, $k_{ij}=1 \Leftrightarrow (iid_{ix},iid_{iy}) \in R_{ij}$, $viid_x$ and $viid_y$ correspond to iid_{ix} and iid_{iy} in SSN_i, respectively, or $k_{ij}=0$ otherwise$\}$, R_{ij} is the jth relationship set from all N_i existing in SSN_i.

[1] For more information about ties see Section 5.4.

[2] A physical social entity is a person, a group of people or an organization existing in the real world. For more information about internet identities see Section 4.1, where also virtual identity was defined (Definition 5).

Note that virtual internet identities merge internet identities from component SSN_i (see Section 4.3), whereas ties from T^M aggregate relationships derived from the component SSN_i, (see Section 5.4).

The example of internet multisystem social network can be *Blogger* (SSN_{12} in Figure 2) that enables to log into the system using either its own user names (IID_1) or external *Gmail* accounts (IID_2). Since both systems have some common user identities, it is possible to merge two system–based networks into one internet multi-system social network.

Internet multisystem social networks can be extracted from two or more system–based networks by merging their internet identity sets, see Section 4.3.

3.2 Other approaches

The social networks of Internet users somewhat differ from the networks existing in the real world where people have in-person contact with each other (see introduction to Chapter 3). Although social networks on the Internet have been already studied in many contexts and various definitions have been created there is no one coherent and recognized model of social networks on the Internet. Furthermore, different researchers name these networks differently. In consequence, they are called: computer-supported social networks (CSSN) [73], online social networks [13, 31, 49], web-based social networks [24], web communities [23], or virtual communities [1], see Table 2 for more details.

In the literature, the term *web communities* was first used to describe the set of web pages that deal with the same topic [23]. Adamic and Adar argue that a web page must be related to the physical individual in order to be treated as a node in the online social network. Thus, they analyze the links between users' homepages and form a virtual community based on this data. Additionally, the equivalent social network can also be created from an email communication system [1]. On the other hand, a computer-supported social network introduced in [73] arises when the computer network connects people or organizations. Finally, Golbeck affirms that a web-based social network must fulfil the following criteria: users must explicitly establish their relationships with others, the system must have explicit support for making connections, and relationships must be visible and browsable [25]. Mainly social networking sites like *LinkedIn* or *MySpace* meet these conditions, see Section 7.4.

Based on the kind of service people use, many examples of the social networks on the Internet can be enumerated. To the most commonly known and studied belong: a set of people who date using an online dating system [6], a group of people who are linked to one another by hyperlinks placed on their homepages [1], customers who buy similar stuffs in the same e-commerce [75], the company staff that communicates with one another via email [15, 36], people who share information by utilizing shared bookmarking systems [53] such as *del.icio.us*.

4 Internet identities – members of the network

In this section, the concept and types of internet identities, which are the nodes of social network on the Internet, is presented. Moreover, the idea of merging few internet identities belonging to one social entity into one virtual internet identity is introduced.

Table 2 Definitions social networks on the Internet.

Name	Definition	Year	Reference	Remark
Computer-supported social networks	When computer network links people as well as machines	1996	[73]	This is a subset of device-supported social networks where people interact using different ways of communication.
Web communities	Web page collection with a shared topic.	1998	[37]	Web page collection is the set of pages that deal with the same topic and link to other pages in this set.
		2000	[32]	These web pages do not have to be associated with individuals.
Virtual communities[a]	A group of people who are linked to one another by hyperlinks placed on their homepages	2003	[1]	The pages are homepages so are associated with a particular person in the real world.
	A group of people who send emails to each other. Relationship between two people exists if there is email communication between them			
Web-based social network	In this kind of network the following conditions need to be met:	2004	[35, 36]	Very detailed and unambiguous definition. Its scope is limited to online networking sites such as *Facebook*, *LinkedIn*, etc.
	Users must explicitly establish their relationships with others, the system must have explicit support for making connections, and relationships must be visible and browsable			
Online social network	The definition of online social network is not really established and different researchers use this term to describe different networks in which people interact with each other by means of different services enable on the Internet (e-mail, forums, blogs, social networking sites, etc) [20, 33, 48, 75]			

[a] Authors use also the phrase web communities

4.1 The concept of internet identities

Each concrete physical individual or a group of people who are the users of internet–based services can possess an internet identity.

Definition 4 *An internet identity iid* is a short digital, verified, authenticable, unambiguous and permanent representation of the physical social entity – a concrete human or group of people, who are aware users of the single internet-based system.

The task of an internet identity is to transfer the physical entity from the real to the virtual world, see Figure 4. This mapping enables to define the connections between social entities based on the connections between their internet identities, see Section 5. Since we are not able to study relationships between physical social entities on the Internet, the only possible social network analysis is the investigation of humans' equivalences, i.e. internet identities.

The concept of internet identities was considered in [38, 62, 63, 69]. Internet identity can also be called online identity [20, 74]. However, it suggests that online identities are restricted only to online, synchronized services and for example email addresses could not be covered by online identities. Thus, authors have decided to use the term "internet identity" as it describes more clearly and in unambiguous way the idea that is behind this expression.

There are seven basic features of internet identities:

a) succinctness
b) authentication
c) uniqueness
d) durability
e) user's awareness
f) correspondence to concrete humans
g) possibility of extraction from internet services.

An internet identity is a short digital representation of physical entity. Hence, e.g. concatenation of the name and the postal address does not fulfil this condition – it is too verbose. Moreover, only authenticable, verifiable users are considered, so they at least have to register to the system. No other action is necessary and the users do not have to use the service any more. For instance, one can register in the e–commerce system, get an internet identity and after that never utilize this account to buy stuffs. Due to lack of relationships, such internet identity would probably be isolated in the social network. Another example that the registration is enough to posses internet identity is the following case: some people send emails to the new, just registered user x and automatically get into relationship with x. Although this new user x may not read these emails and not send any, x possesses own internet identity (the registered email address) and even some relationships with the email senders; everything with no x's involvement since registration.

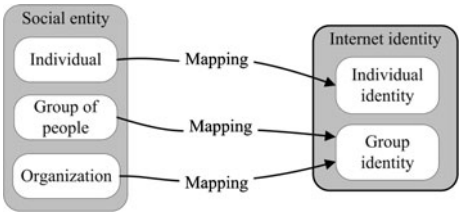

Figure 4 Mapping of social entities into internet identities.

The registration to the service must be done knowingly. Thus, users created by the system administrator should not be considered as the members of the social network unless they are aware of their registration. It may happen that fulfilling of this requirement is hardly to achieve and we would need to assume, especially while automatic processing, that all registered accounts are valid internet identities.

Uniqueness of *iid* has to be guaranteed by the system itself. There should not be two same email addresses on the Internet or two identical user names in the blogging system.

Furthermore, the internet identity must not be temporary. For instance, it cannot be dedicated only to one single user visit in the system.

Some typical examples of internet identities can be enumerated:

- Email address,
- Login, identifier, nick or user name in a specialized system. In such case, *iid* should usually be a tuple (login, system):

  – Registered user name in a social networking site (*Facebook*, *LinkedIn*, *Orkut*, *MySpace*, *Friendster*, *Classmates*), see Section 7.4,
  – Registered user name in an online blogging system (*WordPress*, *Blogger, Twitter*)
  – Instant messenger or VoIP communicator nick (*Skype*, *ICQ*, *Windows Live Messenger*, *AIM*, *Yahoo! Messenger*, *GTalk*)
  – Login to multimedia sharing systems (*Flickr*, *YouTube*)
  – Login to social services like social bookmarking (*del.icio.us*), social travel network (*TripUp*), social searching (*Technorati*)
  – Account in an e-commerce (*Amazon*, *iTunes Store*)
  – User name in auction system (*eBay*)
  – Login to a web-based financial service (*PayPal*, *WebMoney*, ebanks, ebrokers)
  – Registered user in a personalized web portal, especially news service (*My Yahoo!*, *CNN*), online journal (*The New York Times*) or online television
  – Account in a specialized service available on the Internet, for example: online library: ACM or IEEE Web Account with access to *ACM Digital Library* or *IEEE Computer Society Digital Library* respectively,
  – X.509 certificates used to authenticate SSL clients while logging into web sites with restricted access,

- URL to the personal home web page,
- Login to a comprehensive identity system (*OpenID*).

There are also some examples that are NOT the internet identities:

- ID of a single web session – corresponds to the activities of humans rather than the social entity itself, it is temporal
- ID of searching session; ditto
- ID assigned to the exchanged objects, e.g. email ID; ditto
- First and second name of an individual published on their personal web page; it is ambiguous
- Temporal ID assigned to an anonymous user in e-commerce, usually used only for one visit; it is neither authenticable nor permanent
- Anonymous commentator of posts in blogging system; ditto
- Company profile published in the web site; it is not a short digital representation
- Postal address published on the contact web page; ditto
- X.509 certificate (or its serial number) issued to an SSL web server; a server is not a aware user

- Authors of scientific papers gathered in the online bibliographical *DBLP* database (http://www.informatik.uni-trier.de/~ley/db/). Although, it contains data about co-authorship and in consequence their mutual relationships, the authors' names do not reflect internet identities. Besides, the authors are not aware user of any internet service.
- Guest account in an internet service, e.g. one "*student*" account common for all anonymous users in an online e-learning system; it does not correspond to a tangible social entity
- Anonymous account to FTP servers; it is not authenticable
- Trial account in an internet service unless the trial period is long enough; it is not permanent
- Accounts transferred from another system by system provider without user awareness – these new users are not aware unless they accept this operation.

Nevertheless, the thing to remember is that people try to be as anonymous as possible on the Internet. This often is the reason why people multiply the number of their internet identities. Additionally, people may want to separate their private and professional activities [18]. As a result, one physical social entity can possess many internet identities in one system. For example user z possesses one account in the blogging system (iid_5) and one email account (iid_4) in Figure 5. All these z's internet identities can be merged into one virtual identity that represents all z's internet identities: virtual *ID z* aggregates iid_4 and iid_5. On the other hand, one internet identity is connected with only one social entity. In other words, the only restriction for the internet identity is that it has to refer to exactly one physical social entity – an individual or a group of people. Note, that one internet identity can capture several types of identities corresponding to various activities of the person. For example, some people use one common email address in both private and professional life (one *iid* of two types) whereas the others utilize separate addresses for both these involvements (two *iid*s for one person). The same may be valid for social networking sites [18].

Definition 5 Virtual internet identities aggregate distributed internet identities existing in different internet-based systems. A virtual internet identity *viid* corresponds to all internet identities *iid* related to the single physical social entity. Simultaneously, each internet identity is related to only one virtual identity.

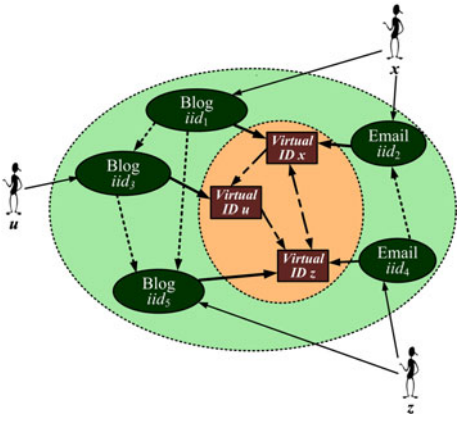

Figure 5 The concept of internet identities merging.

The virtual internet identity is an important concept that enables to analyse the individuals' connections in a more comprehensive and coherent way. Merging all internet identities of a given user into one virtual identity facilitates to represent the whole knowledge about user and his neighbourhoods in networks from which the particular internet identities come from. This concept will be essential for future network analysis as it is very important for researchers not to analyse the behaviour of users in a single system but rather their actions in the whole Web, see also Section 4.3. Additionally, we can create decentralized online networking sites based on virtual identity realized e.g. via FOAF [76]. In such sites security would be managed locally what can make people more trusting.

Note that some users of internet services may correspond to the same social entity in the real world. In some cases, we are able to identify that two different internet *ID*s belong to one physical entity, e.g. based on the data provided by users in their registration forms. Then, we can join two internet *IDs* into one virtual *ID*. The consequence of this kind of merging is the removal of the data about the reciprocal communication between the identities that are merged into one account. Note that this internal communication usually results from the way in which people organize their contacts with others. For example, one can posses two different email accounts – one for private communication and one for professional but emails sent to the private account are usually forwarded to the company mailbox. The similar situation can also occur when the person makes mistakes while using the specific internet service, e.g. one registers to the system many times because permanently forgets the password or login.

In practice, it is usually difficult to merge internet identities related to the same person in an automatic way. However, there are some specialized systems like *OpenID* (see Section 4.3) or *eBuddy* (see Section 7.14) that enable to achieve it with the assistance of users themselves. Additionally, some preliminary research has been carried out in the field of automatic identification of people within social networking sites [43].

4.2 Individual and group internet identity

There are either individual or group internet identities, Figure 4. An individual internet identity belongs to an individual – a single person, whereas a group identity corresponds to a group of people, e.g. family that uses only one login to the blog or to an organization – all employees of the service department use one common email account *service@company.com* to answer customers' requests.

Group identities can be identified by content analysis. If we study the signatures in the emails and we recognize more than one name there then it would mean that more than one person sends these emails. Moreover, sometimes the name of the internet identity can be directly matched with the name of the company, its department or a position.

The interaction between group identities reflects the relations between two groups of people, e.g. two companies, two departments within one organization or two families. In contrary to the individual identities, the group identities are not restricted by social limits of single humans. According to Dunbar's studies, the maximum number of steady relationships that one can effectively maintain is about 150; it is also called the Dunbar's number [30].

Furthermore, the behaviour of people represented by group identities seems to be more stable over time than individual ones, e.g. when an individual goes on leave then the account is usually not used during this time whereas in the case of group identity even if some members are currently not available then the others take these users' duties over. Of course, it depends on the number of people who use this account as well as the type of the group

identity. Probably, the greater the number of real, social entities related to a single group identity the more stable the behaviour of this identity is. For instance, the general company email account used to contact with its clients is likely to be steadier than the identity used by a single family.

Several different types of individual as well as group identities can be enumerated. The former can be:

- private identity, e.g. instant messenger nick to private account, private email address;
- professional identity;
- activity/interest–based identity – the login to the fan club site;
- consumer identity – login to the customer account to the web site of the telecommunication company or e–commerce.

The following types of group identities can be distinguished:

- interest–based identity – special interest groups;
- family–based identity, e.g. a wife and husband can use the same account in the e-banking system;
- task–based – the common account for the project team;
- position–based identity – many people who occupy the same position share the common account, e.g. all PhD students at the small university use the same login to the academic intranet;
- company–based identity, e.g. the homepage where the company provides the information about itself;
- unit–based identity, e.g. the email address of the individual department in the company.

4.3 Integration of internet identities

As it was presented in Sections 4.1 and 4.2, various kinds of internet identities can be distinguished. Nevertheless, nowadays, it becomes more and more popular to merge two or more internet identities into single one in order to facilitate people the access to different services with only one login. Thus, the single sign-on concept (SSO) extends also to the Internet. It is achieved by internal integration of two or more services delivered by a single provider or even the cooperation between independent providers.

The example of such integration can be found within Google services. The single email address enables the user to login into both the blog service – *Blogger* and the email service – *Gmail*. Of course, it is also possible for a person who does not use *Gmail* service to maintain a separate blog account.

Another integration system – *OpenID* allows to create the single common account that facilitates to login to nearly ten-thousand websites with this identity. It eliminates the necessity for creation of multiple usernames across different websites [64]. *OpenID* concept is used among others within *FOAF* format to identify internet users, see Section 7.13.

Yet another example can be *eBuddy* (see Section 7.14) that is a free web–based messenger. This system supports the communication via many other services including *Windows Live Messenger*, *Yahoo*, *MySpace*, *Google Talk* (*GTalk*). Hence, it integrates many internet identities derived from separate systems into one *eBuddy ID*.

Generally, two or more social networks can be integrated based on matching and merging the internet identities existing within all of them. To achieve it, we ought to possess or gain

the knowledge about real users and their internet identities within merged networks that are being integrated. For instance, if two system–based social networks, e.g. VoIP–based social network (1) and the network derived from personal homepages (2) are supposed to be merged then for each social entity the set of the internet identities that a given person possesses in both networks need to be identified. User *a* has both the homepage address and login to the *Skype* system, whereas user *u* is only *Skype* user, Figure 6. We are able to discover internet identities of the same users in two networks using our external knowledge (e.g. data from the paper contacts), matching mechanisms (e.g. by email address) as well as information provided directly by network members – users of the VoIP network can deliver URLs to their homepages at registration to this system or publish their account name to the VoIP system on their homepages. Additionally, the relationships between users from both systems can be utilized in the final, integrated social network (*ISN*) as (1–2) in Figure 6. The integration can provide additional extension possibilities for the merged networks. For example, users *a* and *b* in the VoIP system can be suggested and encouraged to communicate with each other based on the hyperlinks connecting their homepages; the thick solid arrow between *a* and *b* in the network 1-2 in Figure 6. The similar mechanism of merging two networks: a telecom social network and an internet-based network can be used by the telecommunication company to create an additional service for its customers: "call your acquaintances you do not talk to". In this case, the recommended people would be extracted from the internet–based social network [37].

The integration can also be performed based on the user profile merging. For example, if two internet identities have in the demographic profile the same name and address then there is a high probability that they both belong to one social entity (see also Section 4.2). The integration of internet identities can be performed in the automatic way with the certain confidence using so called *personally identifiable information* [43]. Even one third of virtual identities can be recognized based purely only on the network topology, see [57] for experimental results on Twitter and Flickr.

Overall, integration of the internet identities may be a new trend within the web service development.

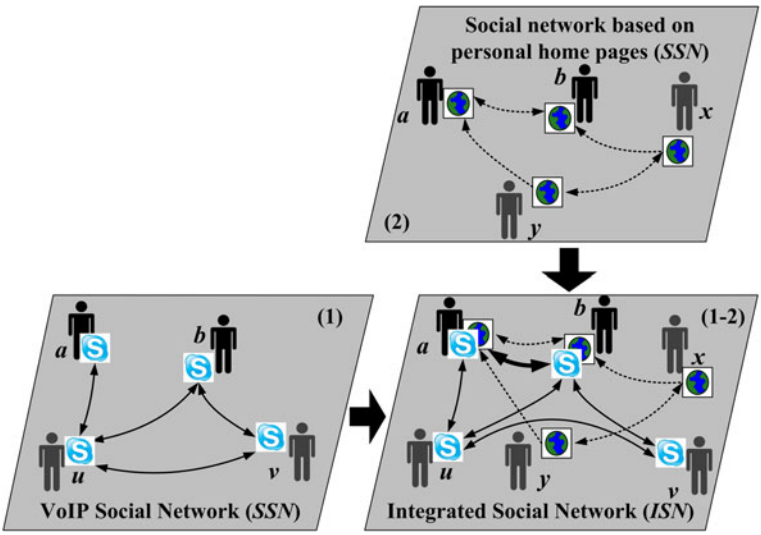

Figure 6 Integration of two system–based social networks by means of internet identity merging.

5 Human relationships on the internet

Apart from internet identities, the second crucial component of every social network is relationships connecting pairs of nodes. The concept and types of both internet relationships and compound ties are presented in this section.

5.1 The concept of social internet-based relationship

Definition 6 *A social internet–based relationship*, called also shortly internet relationships, in the homogenous social network $HSN=(IID,R)$ is the connection $r \in R$ from one internet identity iid_i to another iid_j.

Note that in the system–based social network $SSN=(IID,T)$, a single tie $t \in T$ may contain up to N different social internet–based relationships (definition 2).

A relationship in the social network reflects their common acquaintance, private or professional relation or even high similarity of their inclinations or activities. The maintenance or even only creation of the relationship usually requires member's trust, commitment, emotion, and dedication of time and effort.

Several significant social features can characterize a human relationship, in particular [72]:

- mutuality,
- durability,
- intensity,
- intentions,
- culture conditionings,
- emotional level,
- strength.

A relationship does not have to be symmetrical, e.g. Tom could be friends with John but John might not see Tom as his friend. Nevertheless, if a relationship is symmetrical then it is usually more durable. Moreover, a relationship may be durable for a certain period; afterwards it could significantly weaken or even diminish. If Tom sent John 20 emails over two weeks, but five years ago, then John would have most probably forgotten Tom by now. On the other hand, John would remember and feel a kind of durable relationship with Bill who has regularly sent John one email every quarter for the last five years. The number of emails is the same in both cases (20) but the latter appears to be much stronger right now. Furthermore, the longer the acquaintance, the more durable in future terms.

The importance of contact intensity and communication features on the strength of the relationship may result from the culture both participants live in. Ten emails sent by the representative of one nation may have greater significance than the same number of emails exchanged between individuals from other, more spontaneous nations. Many phone calls made late at night or in time off are the sign of more private and thus stronger relationship than the same calls performed in regular working hours.

The strength of a relationship can also depend on the kind of communication or mutual activity based on which the relation is created. The meeting of commentators of the same blog or even hyperlinks between homepages generally connect people much lesser than the co–authorship of a scientific paper.

Some unusual factors may also be the sign of stronger relationships. An intensive correspondence in Polish is the evidence for stronger relationship between foreigners in Japan rather than the same communication in Japanese between natives. Of course the opposite meaning would be true but in Poland.

In some environments like the worldwide Internet, that is multicultural in its nature, the detection of some differences appears to be very complicated. Moreover, some features of human relationships may either require complicated content processing like extraction of the emotion level or even be very hard to discover like intentions.

Note that definition 6 assumes that a single relationship binds only two internet identities. In more general approach, we can use hyperedges and a hypergraph as the representation of the social network. A hyperedge connects any number of network nodes, at least two. This can be useful especially in the case of relationships derived from common activities or interest as well as based on profile matching. Comments on a single blog post involve all participating commentators; single interest can be simultaneously shared by many people; many members can have profiles similar to each other, etc.

5.2 Types of internet relationships

The relationships existing on the Internet can be classified in many different ways and based on different characteristics (Figure 7):

- active subject that is responsible for creation of new relationships (user, system, user and system),
- awareness of the users that they are involved in relationships,
- mutuality of the connection between users (asymmetrical, symmetrical, reflexive),
- general relationship sources (external or virtual world),
- data type used by the system for relationship creation (direct interaction, common activity, user profiles, none if the relationship is created by user),
- nature of relationships (professional, family, friendship, acquaintance, common interest, etc.),
- visibility of relationships for the users (fully or partly visible, invisible).
- directness of relationship grounds (direct, quasi-direct, indirect). This is an original concept developed by authors and is described in Section 5.3

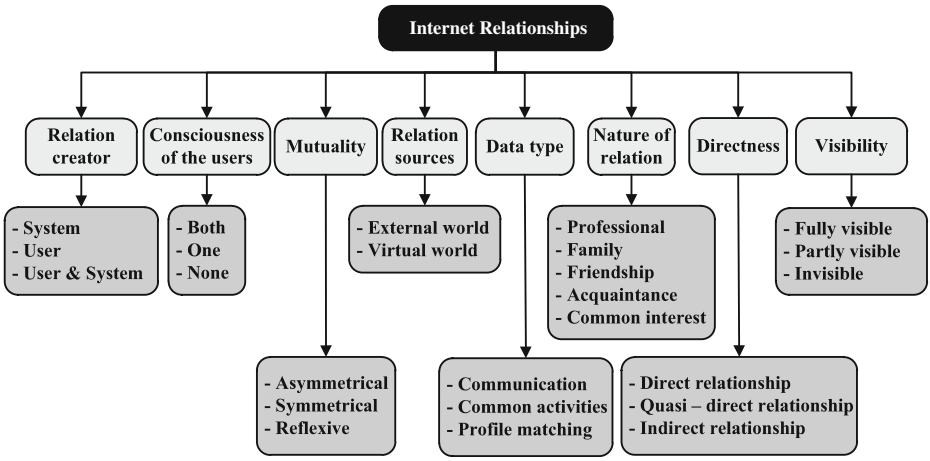

Figure 7 Taxonomy of internet relationships.

5.2.1 Active subject creating relationship

In the first classification, the relationships can be divided in three main types: created by the user, established by the system and the mixture of the two previously enumerated. In the first type, user *x* can for example set up a relation with person *y* by adding *y*'s email address to *x*'s private contact list or link to *y*'s homepage at *x*'s private page. The new connection starts up also when two people exchange emails or one of the user sends invitation to another within the social networking site (like *Facebook*, *MySpace*, or *LinkedIn*) and the other person accepts this invitation. Nevertheless, the relations can be initiated and created also by the system itself, for example, when the profile matching is performed. The last but not least situation is when both the system and user are involved in the process of relationship creation. For instance, when the system recommends other users to the specific one, then it admittedly initiates the relation but the user has to confirm that is really interested in such a new relation by approving the suggestion generated by the system. Only when the user accepts the recommendation the connection is created. Analysing these three types of relations the correlation between connection origin and such characteristics as relation durability or intensity can be investigated. This classification is useful when one is interested in origins of the relationships as this enables to identify users who are very active in creating new relations. The shortcoming of this approach is that in a specific system there can be only one type of relationship, e.g. in email system the only entity that creates relationship is human being so the analysis regarding the origin of relation and its influence on the further connection evolution cannot be performed.

5.2.2 Awareness of users involved in relationship

When the awareness of the users that are involved in the relationship is considered then three kinds of connections can be distinguished. The first type occurs, when both internet identities participating in the relation are aware of this fact, e.g. two users communicating with the instant messenger or exchanging emails. The second situation happens when only one side of the relationship is aware. The example for this can be adding by single user *x* another person *y*'s email address to *x*'s private contact list or link to *y*'s homepage at *x*'s web page. Person *y* is usually not aware of these activities of user *x*. In the third type, we have relationships in which none of the participants is aware of the connection existing between them, e.g. when the relationship is created by the system based on the profile matching; it is also called demographic filtering. This classification can be useful when the analyst wants to find out what types of relations dominate in a specific network and how this distribution influences a level of activity within this network. If the relations, in which both user are aware of it, constitute the majority of the connections the system has greater chance to survive than in the case when other relations dominate. Analysing the dynamics of relations and when and how they change their type from unaware to aware ones can be a very interesting topic in relations dynamics. The durability and intensity of relations where both users are aware of them can be analysed depending on the level of awareness of these relations at the beginning, e.g. the important question is which aware relations are more durable – those in which both users are aware of the relation from the beginning or maybe those where at the beginning users were not aware of the relationship but now they are? However, the analysis of comparison of the three levels of "relationship awareness" and their influence on relation durability and intensity do not have sense as the connections in which the users are not aware of do not have any dynamics.

5.2.3 Direction of relationship

The next feature of social relationships is the direction and mutuality of the connection between users, Figure 8.

The relationship can be asymmetrical, i.e. internet identity iid_x is in the relationship with internet identity iid_y but there is no reverse connection from iid_y to iid_x, Figure 8. The example of such relation can be if user x adds user's y blog to the favourites but user y does not do the same. The symmetrical relationship exists when there is a mutual communication between users or when people share common activities, e.g. comment the same photo in the multimedia sharing system such as *YouTube*. Due to social and collective profile of social networks all reflexive relationships are usually excluded from consideration.

This classification is the most common one as there is always data that enables to determine the direction of the relation and this is its main advantage. The investigation of the relations dynamics depending on if they are asymmetrical or symmetrical can be performed. Moreover, a very interesting issue is to what extend and when the asymmetrical relations change into symmetrical and also the other way around.

5.2.4 Source of relationship

The connection between two internet identities can be also investigated based on its source, i.e. where does it origin from? The acquaintance can come from the external world, e.g. two network members know each other personally and they have exchanged their email addresses. When they start sending emails to one another, then the relationship is set up in the virtual world. However, a relationship can also exist only in the virtual world. This situation appears, e.g. when one user sends the invitation to another one within the social networking site such as *MySpace* and additionally these users previously did not know each other in the real world.

This distinction is a very interesting one as one can analyse how the source of relation influences its future evolution in terms of its strength and dynamics. However, the problem is that there is no information about relations originating from real world regarding their history before they entered virtual world. Thus, the further analysis can be biased by this lack of data.

5.2.5 Data used to create relationship

Another approach to relations classification is to split them according to the type of data used by the system for relationship creation. In consequence, an acquaintance can be created based on the data about mutual communication (email exchange), common activities (commenting the same multimedia objects, using the same commercial internet service), data derived from the profile matching, or data from users' contact lists (e.g. contact lists from instant messengers). Dividing relationships into classes depending on the type of data used to create these relations add additional granularity level to the analysis. It enables to look at connections characteristics and investigate the correlation between these features and type of data used to relationship creation. The shortcoming of this method is that although

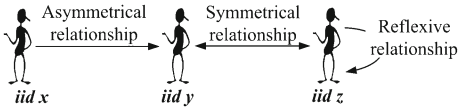

Figure 8 The direction (mutuality) of relationships [16].

the relationship itself could be created based on one type of data, after that the users start using different means of communication so this criterion has limited usefulness in network analysis. For example the relation was created because two users commented the same photo but after that they just exchange emails. The single activity of commenting photo is not very informative in the context of social network analysis.

5.2.6 Nature of relationship

One of the most interesting classifications is the one according to the nature of relationships. Hence, among many types of relationships the following can be distinguished: professional, family, friendship, acquaintance, common interest, customer relationship (online consulting, e-learning, usage of specific internet service or its features), etc. Nevertheless, the process of specifying the character of the relation is a very complex task because it is hard to identify in the virtual world what kind of relation exists between two users unless they state openly the profile of their connection. Another method that can serve to recognize the character of the relationships is the investigation of the parameters of the communication between two users (in particular its time and frequency) or common activities. When more than one "nature of the relationship" between two people can be defined, then we talk about more than one relationship and the all of them together create a tie, see Section 5.4.

As in the previous cases this classification can be used to identify common features of relationships of a given nature. In addition, it enables to investigate how many different relationships exist between two actors and the dynamics of relationships of a given nature. However, this criterion is not objective and only people involved in the relation can subjectively determine what the nature of relation is for them. This causes that although this distinction is very interesting from psychological point of view, it is not applicable from network analysis perspective.

5.2.7 Visibility of relationship

The relationships can be also classified based on their visibility for other users. It especially regards the social networking sites like *Facebook* or *MySpace* where people can directly define who can browse their profiles and relations. The number and specification of the visibility levels depend on the system, e.g. in *Friendster*, users decide whether their relationships can be viewed either only by the nearest friends, also by friends of a friend or by the whole community. This criterion can be used to assess the level of trust between different users. Moreover, this also may be used to define the number of friends that one can maintain. The shortcoming of this classification is that user decides about the visibility of the relations that he/she is aware of so the rest of the relations are not considered within this classification. Because user decides about the relation visibility it causes that this criterion as the previous one is subjective.

5.2.8 Discussion

Each of the abovementioned classifications provides some additional opportunities to analyse relations characteristics and their dynamics. Information about to which classes the relation belongs to is an extra input for advanced network analysis.

The disadvantage of the presented approach is the assumption of the availability of data required to classify a given relationship. However, once the data is gathered the classification

provides a comprehensive and coherent view on each relationship that can be identified on the Internet and this completeness of description is a big advantage of such approach.

Moreover, the limitation of each of the enumerated criterion is that it analyses only one feature of the relationship. To fully describe a given relationship one needs to assign to it one value for each of the defined above features. As a result, we obtain eight dimensions of the relationship.

For example let us assume that two people work in the same company and have exchanged their email addresses. The relationship is created when one person adds email address of another person to the contact list. Such relationship can be described in the following way:

- relation creator – *user*,
- consciousness of the users – *one user* who adds the email address to the contact list is aware of the fact that relation has been created,
- mutuality – *asymmetrical*,
- relation source – *external world* as the people met previously and have exchanged their email addresses,
- data type – *none* as the relationship is created by user,
- nature of relation – *professional*,
- directness – *quasi-direct relationship* – there is a meeting object called contact list,
- visibility – *invisible* – only owner of the contact list can see it.

Each of the considered taxonomies is applicable depending on the type of analysis that is performed. However, some of the criteria are subjective and depend on user's perception (e.g. nature of relation or its visibility) what make them hard to apply in unbiased and objective manner. Other provides knowledge just about subset of relations, e.g. the criterion: "data type used to create relationship" can be only applied in the situation when the system has created the relationship. While considering relationship source one faces the missing data about the relationships originating from real world. The "relation creator" criterion does not provide information about how the relationship was created but only by who/what it was done. Similar situation is in the case of both "consciousness of the user" and "relationship mutuality" as they inform us only about user awareness of the relation and relationship direction respectively but not about how it was created.

The identified shortcomings of existing classifications show the need for one comprehensive classification that enables to classify in an objective manner all relations that can be identified between users in the Internet. The proposed by authors classification presented in Section 5.3 encompasses information about relation creator, awareness of relation, its mutuality and data type used to relation creation. Relationship source, its nature and visibility are not reflected in the proposed classification as they are subject to data availability (e.g. in the Internet one cannot assume that prior information that is not stored about relationship is available) or vary depending on people perception (one can perceive friendship and trust differently than other person).

5.3 Directness of internet relationship grounds

One of the taxonomies presented in the previous section was the classification of relationships with respect to their directness. It is of great importance for the future research in the area of social networks on the Internet as analysis of relationships from this perspective provides more information about the users and their interactions than other enumerated above taxonomies.

We can distinguish three kinds of relationships: direct, quasi–direct, and indirect, Figure 7.

Social entities related to the internet identities can be more or less aware of the relationships they are involved in and this partly depends on the basis where relationships are derived from. For that reason, three kinds of internet relationships can be enumerated:

- *Direct relationship* – is a relation that connects two internet identities with a direct connector, Figure 9. The direct connector is an object that is addressed to the specific internet identity and is usually related to the specific feature (communication, activity) existing in the system. For example, an internet identity establishes and supports a direct relationship while sending an email to another internet identity. Thus, the direct connector can be derived from an email, a phone call (VoIP), message sent by means of instant messengers, hyperlink binding one home web page with another one, an item in the somebody's contact list, a connection in the social networking site (see Section 7.4), etc.

  Quasi–direct relationship – two internet identities are in the relationship but it is not required that they maintain the relationship themselves, e.g. people who comment on the same blog or participate in the common business meeting. There is always a meeting object, which serves as the communication medium between users, Figure 10. The roles of both internet identities, which are in this kind of relationship, in relation to the meeting object can be either the same or different.

  Quasi-direct relation with equal roles r_{xy} means that internet identity iid_x and iid_y meet each other through the meeting object and their role in relation to this object is the same. In other words, they participate in common activity related to the certain object with the same role a, e.g. two users comment the same picture, both of them add the same object to their favourites or both use the same tags as metadata to describe their photos, Figure 11a.

  Quasi-direct relation with different roles r_{xy}^{ab}, r_{yx}^{ba} – is the relation between two internet identities u_x and u_y that are connected through the meeting object (multimedia object or their additional feature like tag) but their roles a and b towards the meeting object are different, e.g. u_x comments a photo (role a – commentator) that was published by u_y (role b – author), Figure 11b. The non zero relation r_{xy}^{ab} entails the non zero relation r_{yx}^{ba}.
- *Indirect relationship* – this kind of relations exists when the internet identity is not aware of the fact that is similar to another internet identity. Two internet identities are connected by indirect link when their profiles are similar, Figure 12. If these relationships are discovered and analyzed in a right manner then such knowledge can be used to change the hidden relationships into direct ones.

It is worth to notice that the direct relationships can be supported and developed by utilizing the knowledge derived from the characteristic of indirect relationships, e.g. the recommendation systems can use the demographic filtering to suggest movies liked by people with the similar taste, i.e. with the similar profile.

It is worth to emphasize that during performing the process of identifying the directness of relationship also the knowledge provided by other enumerated in Section 5.2 taxonomies is discovered.

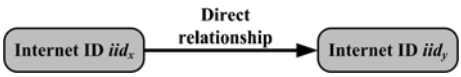

Figure 9 The direct relationship in the social network on the Internet.

Figure 10 The quasi direct relationship in the social network on the Internet.

Having defined directness of connections also some other characteristics can be analyzed: (i) the type of the active subject that is responsible for creation of new relationships, (ii) awareness of the users that they are involved in relationships with others, (iii) mutuality of the connection between users (asymmetrical, symmetrical, reflexive), (iv) the data type used by the system for relationship creation. It means altogether that directness of internet relationships grounds encompass more information than any other presented classifications.

5.4 Ties

The issue that is tightly connected with the concept of relationship is a tie. A tie is the set of all relationships that exist between two internet identities. In other words, a tie between two users aggregates all types of the relations that exist between these two persons.

Similarly to virtual internet identity, a tie provides more comprehensive and coherent view on users' interactions. Combining different internet relationships existing between two users into one tie, an extra, before unavailable knowledge can be discovered. The analysis of networks where more than one type of relationship exist were performed for example in [33, 34].

The types of relationships, which create a single tie, can reflect different communication channels used to exchange information. For instance, two users who send emails to each other use *SMS* and *VoIP* services for mutual communication maintain three types of relationships. In such case, the tie that exists between them consists of three separate relations, Figure 13. An analogous situation appears when the complex character of mutual relationship is analyzed. Two people can be in the relationship of a friendship and in the same time co–work together. In consequence, they maintain two types of connections and the set of them is called a tie.

Note that in a homogenous social network *HSN* a tie is the synonym of the relationship, because there is allowed only one kind of the relationships in *HSN*.

The different types of relationships (e.g. friendship, family, professional, etc.) can be grouped into layers. A layer of the social network is, in fact, the single *HSN*. During the research that the authors conducted on Flickr dataset [35] eleven types of relations were identified: relations created based on contact lists, tags used by more than one user, user groups, multimedia objects (pictures) added by users to their favourites, and opinions about

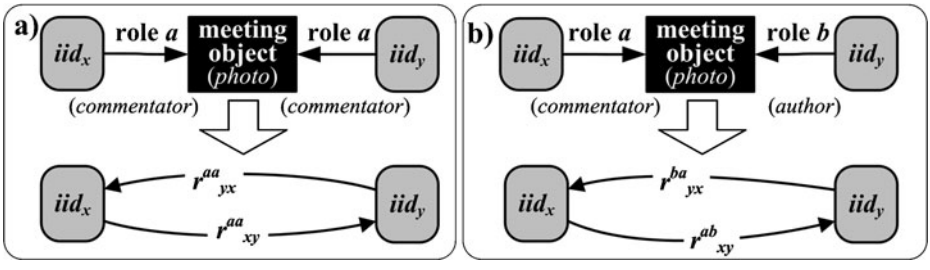

Figure 11 The object-based relation with equal roles: *commentator* (**a**), and different roles: *commentator* and *author* (**b**).

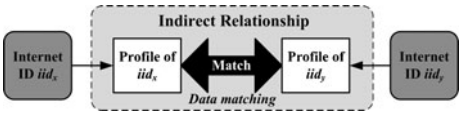

Figure 12 The indirect relationship in the social network on the Internet.

pictures created by others. Relations based on contact lists represent direct intentional relations. Tag–based, group–based, favourite–favourite, and opinion–opinion relations are instances of object–based relations with equal roles, whereas favourite–author, author–favourite, opinion–author, and author–opinion are object–based relations with different roles. All these relations were the basis to create nine separate layers in the social network, Figure 14. In consequence, each of the layers creates the separate *HSN*.

6 Taxonomy of social networks on the internet

Social networks on the Internet can be divided into several groups in terms of different criteria. They can be: dedicated SN (e.g. dating or business networks, networks of friends, graduates, fun clubs), indirect SN (instant messengers, address books, emails), common activities SN (e.g. co–authors of scientific papers, co–organizers of events), hyperlink networks (links between homepages), etc.

To put these different kinds of networks in order, they can be classified with respect to the following criteria, Figure 15:

- the purpose of the internet service,
- the character of the relationship that connects two *IID*s,
- the type of the internet identities that build the social network,
- the type of the communication channel between members that serves to exchange resources; usually the communication channel (internet service) also determines two other criteria: real time / non real time and access openness,

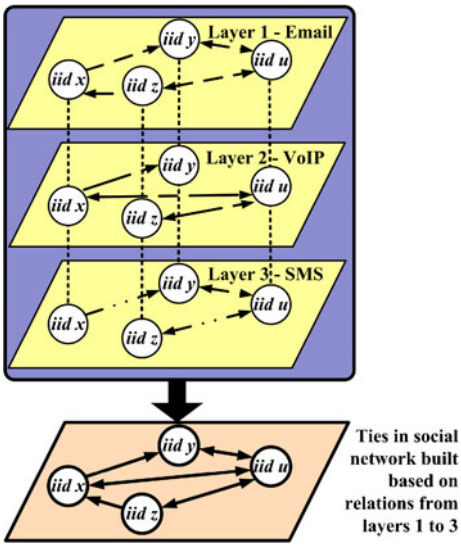

Figure 13 The tie concept in the social network.

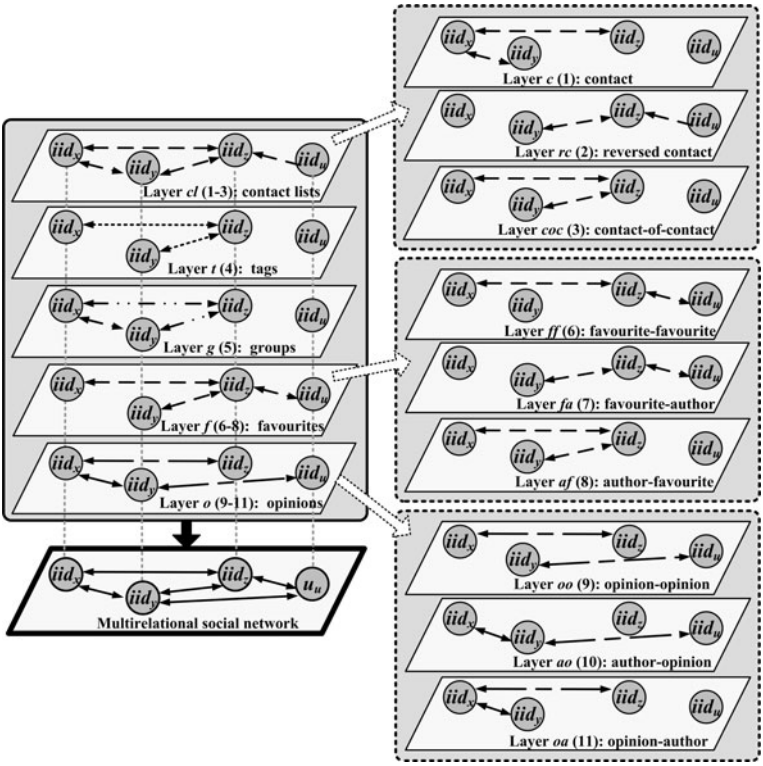

Figure 14 The relation layers within the *Flickr* system.

- real time or non real time networks,
- the type of the access to the network, (open/restricted access),
- the level of the member awareness of their relationship with other members,
- dedicated– or common–service–based social networks.

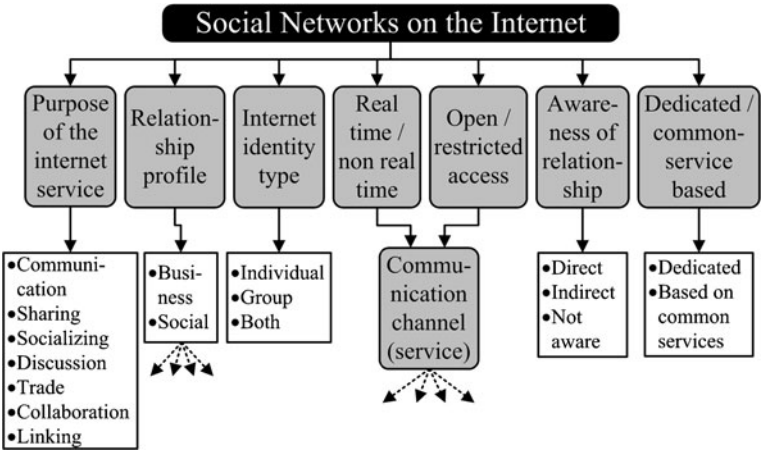

Figure 15 General taxonomy of social networks on the Internet.

The first classification can be made according to the purpose of the internet service and social networks can be extracted because of the following reasons:

- For communication: email, instant messengers, IRC, complex communication systems;
- For sharing user achievements: blog services, multimedia sharing systems, social bookmarking, cataloguing and citation, knowledge sharing systems, clipping systems, social search engines, sharing specific information like own genealogy (*Geni.com*, *MyHeritage*, *WebBiographies*), writing (*Helium*), treatment and symptoms (*PatientsLikeMe*), diet programs (*SparkPeople*), etc.;
- For socializing: social networking sites;
- For discussion: internet forums;
- For trade: auction systems, ecommerce, price comparison services;
- For collaboration: virtual worlds, multiplayer online games, collaborative authoring systems and wikis, clipping systems, web-based e-learning systems, including learning of foreign languages (*italki.com*, *Livemocha*), collaboration within a specific domain, e.g. collaboration between scientists (*SciSpace.net*, *Epernicus*, *ResearchGate*), travellers (*TravBuddy.com*, *Travellerspoint*, *WAYN*), activists (*Care2*), charity (*SocialVibe*), etc.;
- For linking with friends and interesting people: social networking sites, homepages, *FOAF*.

In the classification based on the relationship profile, two essential kinds of relations can be distinguished: business and social connections, Figure 16. The former are included in professional social networks that consist of people who are linked with each other due to common professional activities but simultaneously they do not share their private lives [19]. For example, the email correspondence between employees within the company can serve to extract the social network of co–workers. Also, people who organize something together, e.g. a conference or another event, create a social network of co–organizers. These people are connected because they work together and their cooperation usually brings some outcome, e.g. publishing a common article or book, organizing a joint conference, etc. On the other hand, the social relationships indicate the connections with emotional background. Relatives are the group of people that are a family; regardless people are not in touch with every member of their family.

Obviously, all classifications of relationships types presented in Sections 5.2 and 5.3 can also be the examples of social networks taxonomies made according to the type of relationships.

Moreover, the classification of social networks can also be made based on the types of internet identities that are the elements of the particular social networks. Overall, three types of such networks exist, i.e. those that consist of (i) only individual identities, (ii) only group identities, or (iii) both of them. The most common are the networks containing the mixture of both types of identities.

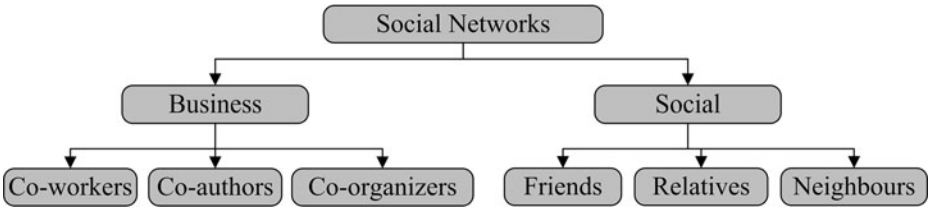

Figure 16 The division of social networks based on the type of the relationship between *IID*s.

Another classification of social networks can be based on the type of the communication channel, i.e. email, instant messengers, VoIP systems, video conferencing, etc., Figure 17.

In general, social networks on the Internet can be divided into non real time and real time networks, Figure 17. The former enable asynchronous communication between two persons or from one person to a group of people [73]. The example can be email system. In contrary to email system, which supports the non real time communication between either two persons or small selected groups of people, the internet forums, blogospheres and multimedia sharing systems like *YouTube* enable all users from the given community to read all messages submitted by every single member of the network. Their functionality is similar to bulletin board from the real world.

Chats, instant messengers, and VoIP systems create the second group of social networks i.e. real time networks. Here, the communication between users is synchronous, for example to take part in an online chat (e.g. *Internet Relay Chat*) the user has to be online. Chats enable to submit messages that will be seen by all people who participate in and who will have opportunity to answer to these messages. The instant messengers (e.g. *ICQ*) serve to exchange information between two persons or limited group of people. The development of the Internet caused that not only text messages can be exchanged, but also voice and video streams. These media are used by VoIP systems, e.g. *Skype* or *Ventrillo*, increasing social presence [73].

There are also some hybrid systems that provide both synchronous and non synchronous communication like auction systems. Online users of such service can observe results of their activities immediately, but they can also be informed about some other activities offline, e.g. via email.

Another classification of social networks can be done based on the type of the access to the social network. The networks can be either with the open public or restricted access. In the former everybody can join them, e.g. *Facebook*, *MySpace*, *ICQ*, etc. while in the latter if one wants to become a member then somebody else who has already been a member must invite this person, e.g. *LinkedIn*. There also exist networks with the restricted access, which are limited only to the people who belong to the specific group or company.

The users of the service can be of different level of awareness of their relationships with other members of the network. In particular, there can exist a visual interface that directly presents relationships of a given user. This is a common case for social networking sites, see Section 7.4. However, people can get into relationship according to their activities, e.g. via common activities, e.g. commenting the same picture in photo publishing site. The relationships derived from such information are not directly visible for the network member, however, one may be indirectly aware of such connection, see also Section 5.3. There are

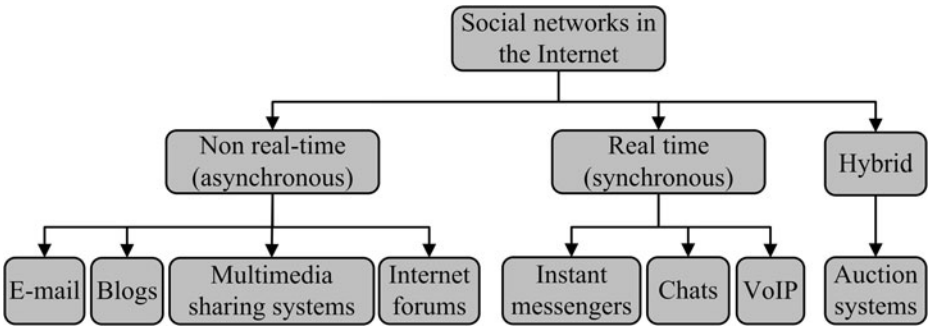

Figure 17 Real time vs. non-real time networks for different services (communication channels).

also social relationships that are hidden for the people and only computer processing can make them visible. The system of hyperlinks between web home pages is an example of such relationships.

Additionally, social networks can be either the crucial or additional feature of the services existing on the Internet. In case of social networking sites, the social network linking its users is essential for most of its functionalities, whereas social network derived from email communication requires specific processing and it is only a potential, additional aspect of the service. As a result, we can enumerate social networks either dedicated or based on common services.

The above proposed classifications are not the only possible ones. However, they pinpoint that there exist many possible taxonomies of the social networks.

7 Categories of social networks on the internet

In order to describe the variety of social networks that exist within the Internet some common examples are presented and compared with one another in this section.

7.1 Electronic mail services

Email systems are the bidirectional and asynchronous way of communication, see Section 6, which enable users, who are in different places and on different schedules, to communicate with one another by exchanging messages [73]. This is the basis to form a social network, in which email addresses represent physical social entities. The relationship in an email–based social network can be derived both from the communication between senders and recipients as well as from the address books maintained by users, see Section 5.3. The registered email addresses and information about communication (logs of SMTP servers) as well as some information about private address books stored on the server can be acquired from separate, distributed mail servers, e.g. *Gmail*, *Yahoo! Mail*, *MSN Hotmail*, *AOL Mail*, etc. On the other hand, many email users utilize own local email transfer agents (MTA) and maintain their address books only on the local storage. Obviously, this data is unavailable for external processing. Address books and communication (logs of exchanged emails) are two main sources to create, analyze and explore email–based human relationships. They can be treated either as the separate layers within the system–based social network *SSN*, see definition 2 and Section 5.4, or as two homogeneous social networks *HSNs*, see definition 1.

Note that email communication is characterized by the high rate of fake or uncertain data. Hence, during analysis of email–based social networks the cleansing process appears to be a very complex and challenging task. It includes especially removal of spam mails and merging many email addresses that can belong to single social entity.

7.2 Instant messengers

The instant messengers (IM) such as *ICQ*, *Skype*, *Windows Live Messenger* (former *MSN Messenger*), *AOL Instant Messenger* (*AIM*), *Yahoo! Messenger*, *Google Talk* (*GTalk*) serve to exchange information between two persons or limited group of people. The rapid development of high speed internet connections caused that not only text messages but also voice and video streams can easily be transferred online. These features are incorporated into many VoIP systems, e.g. *Skype* or *Ventrillo*, increasing social presence [73]. Nowadays, most of the instant messengers support also other kinds of communication channels. However,

their primary goals are pretty precisely defined – exchange textual information. In contrary to email systems, an instant messenger provides the easier way of collaboration because it offers a real time communication. Additionally, it is usually visible for the user whether other people from their contact list are recently available or not because there is the possibility to set the recent user status, e.g. *online*, *away*, *not available*. Since most data related to individual users is stored locally on their computers, the acquisition of communication data necessary to build the social network from instant messengers may be very difficult. Nevertheless, some IM operators provide the opportunity to transfer and retain some local data on the central server.

7.3 Blogs services

The blogs services like *WordPress*, *Blogger*, *LiveJournal* or *Windows Live Spaces* are not only the online diaries but they can also be treated as the system–based social network [28, 50]. In this case the login to the system is the single internet identity. The relation between two *iid*s can be created based on the list of favourites, tags commonly used, or comments made on the blog. All of these connections are quasi–direct relationships, see Section 5.3. In the first case, the favourite–favourite relationship denotes the connection between two people who added the same blog to their favourites, whereas relationships favourite–author and author–favourite reflect the acquaintance between person who has added the blog to favourites and the author of this blog. Similar distinction can be made for relationships based on comments added to posts on a particular blog. Tag-based relationship exists if there are two users that have used the same tags to describe the content of their blogs. Data gathered about users and their activities within blogspace enable to analyse their evolution [45]. Quite similar functionalities and in consequence relationship types provide microblog services like *Twitter* [32, 46, 57], except the limited length of the messages published, e.g. up to 160 characters.

7.4 Social networking sites

In the past few years, the popularity of social networking sites (SNS) [7, 14, 19, 27, 59] has rapidly increased. They can also be called virtual communities, social network services [14], social network systems on the Internet [13], online social networks [27, 31, 43], online networking sites [27], social web sites [39], social networking portals [59], social network sites [7], or Internet social networking sites [68]. Their main goal is to create, maintain and present social relationships to their users as well as match them with each other. To achieve it, they make use of some additional communication services like emails, chats, instant messaging. Recently, the concept of social networking sites together with publishing and blog services is commonly named social networking [19] and the common term for the systems is social websites [41].

Typical examples of social networking sites are: *Facebook* [13, 18], *Friendster* [6], *Orkut*, *Tribe*, *Ecademy*, *MySpace* [12] and *LinkedIn* [17, 49], which are created and maintained by commercial companies. The main features of social networking sites are: self-expression (maintenance of personal profiles), including presentation of personal achievements, striking up relationships with others and mutual communication. There are several ways of communication between users within these online networking sites, which vary depending on the functionality of the portal: email, chat, forum, blog, comments, testimonials, photo/movie album, etc. Overall, the more communication channels are served by the network, the better. This provides greater opportunity to create some new and maintain the existing relationships within the system.

In order to understand the concept of SNS, their main functionalities are described. Usually, at the registration stage, each user should fill in their personal profile (Figure 18), that contains for example their demographic data, information about their hobbies, professional experience and general profile of people that they are interested in.

Afterwards, user x sends or receives invitations from other users of the network. If either user x replies to invitations or other participants reply to user x's invitations, then the relationship is established, Figure 18. However, not only the initiation of the relationship is important, but also its maintenance is one of the crucial parts of every SNS.

Many, different types of social networking sites can be distinguished. One of the interesting classification criterions is the profile of users the networking site is targeted at. Based on this, the following social networking sites can be enumerated:

- general – *Facebook*, *Friendster*, *Orkut*, and many local services usually restricted to the single language;
- dating – *Yahoo! Personals*, *OkCupid*, *Fubar*, *Match.com*, *eHarmony*, *Plentyoffish*, *Zoosk*, *Christian Mingle*, *JDate*;
- school alumni – *Classmates.com*, *Friends Reunited*, *Nasza Klasa* and recent students – *College Tonight*, *StudiVZ*;
- professionals – *LinkedIn*;
- scientists – *SciSpace.net*, *Epernicus*, *ResearchGate*;
- artists – *DeviantArt*, *quarterlife*, *Taltopia*;
- activists – *Care2*, *WiserEarth*;
- interested in politics – *dol2day*;
- fantasy fans – *Elftown*;
- teenagers – *Piczo*, *Faces.com*, *Habbo*;
- mobile communities – *itsmy*, *MocoSpace*, *mobikade*;
- religious people – *MyChurch*, *Xt3*, *Muxlim*;
- businessmen – *Talkbiznow*, *XING*;
- customers – *Yelp, Inc.*, *Epinions.com*;
- competitive endurance athletes – *athlinks*.

The deeper analyses of this kind of systems were made, e.g. in [7] and [59]. In the former, the definition and history of SNS was presented whereas in the latter, the authors classified SNSs according to the following criteria: whether they are registration– or connection–based; whether

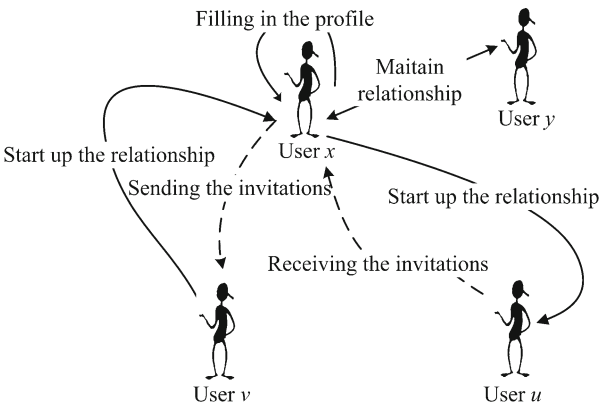

Figure 18 Main functions of the social networking site related to relationship maintenance.

user profiles are socially (*Friendster*) or professionally oriented (LinkedIn) [19] as well as if explicit relationships can be defined; and if the sites are not–for–profit or profit–based. Some of these sites, namely *Orkut*, *Friendster*, *Friendzy*, *Tribe*, *Tickle*, *LinkedIn*, *Spoke*, *Ecademy*, *Ryze*, *Meetup*, were compared in [59].

7.5 Multimedia sharing systems

Multimedia sharing or publishing systems (e.g. *Flickr* [57], *YouTube*, *Photobucket*, *ImageShack*, *Break.com*, *Google Video*, *Metacafe*, *OneWorldTV*) enable a user to upload and manage their own multimedia contents such as photos, videos, animations commonly called multimedia objects (MOs). Each of the multimedia objects can be tagged by the author. In other words, a user can describe their MOs with one or more short phrases that usually denote the content of this element. These tags used by the members can be the basis to create a social network based on tagging, in which a relationship between two members exists if both of them have used at least one common tag to describe their multimedia objects. Simultaneously, users have the opportunity to interact, collaborate and influence one another in different ways, including video-based interactions. Hence, they can not only tag the items they have published but also comment MOs added by others, include them to their favourites, etc. Additionally, users have the opportunity to set up new, direct relationships with other system users.

People who cooperate with one another or share common activities via publishing system can be seen as a specific social community. The members of this community, represented by their internet identities, can be related either directly or indirectly, see Section 5.3. Direct relationships are derived from connections explicitly provided by users who, for example, place other users into their contact lists. Nevertheless, two or more internet identities can also be related indirectly through an external object like a group or tag they share or an item they commonly comment. Users even do not need to be aware of the indirect relationships they are involved in.

The *Flickr* system is an example of such a multimedia sharing system. In this photo sharing system, nine miscellaneous relationship layers can be identified from the data about user activities, i.e. contact lists, tags, groups of items, favourite pictures, and comments to photos [35]. Some of them like favourites and opinions were split into three separate layers, e.g. author–commentator, commentator–author, commentator–commentator, Figure 14. *Flickr* with its layers is a typical system–based social network *SSN*, where a single set *IID* of internet identities exists and nine different types of relationships can be distinguished: R^c – contact based, R^t – tag-based, R^g – group-based, R^{ff} – favourite–favourite, R^{fa} – favourite–author, R^{af} – author–favourite, R^{oo} – opinion-opinion, R^{oa} – opinion–author, and R^{ao} – author–opinion, see definition 2. Similar relationships can be recognized in every multimedia sharing system.

7.6 Auction systems

The main goal of the auction systems is to enable people to sell and buy different stuffs to and from other users. The examples of such systems are *eBay* [5] or *OnSale*, in which people as well as businesses can buy and sell their goods and services worldwide. Obviously, every person who wants to use such a system must register with a unique name that becomes the user internet identity *IID*. After the log in to the system, the members can create new auctions and sell things as well as buy different stuffs so the natural relationships between buyers and sellers emerge. Nevertheless, this is not the only type of relations that can be identified in the systems of this kind. Potential buyers can ask the seller a question referring

products they offer using additional system functionalities. Usually, sellers may remove some bids provided by unreliable users establishing in this way a kind of negative relationship. Moreover, users have the possibility to directly invite some selected members to participate in the auctions they manage. Once the auction is completed, both the winner and seller can comment the quality of the service. Additionally, indirect relationships between buyers or sellers can be extracted due to similar stuffs bought or sold, see Section 5.3. The auction systems with their functionality (especially ratings) provide the opportunity to analyze not only the existence of the relationships but also their intensity and dynamics [66]. Social network analysis, in turn, can be used to support crucial issue in auction systems – trust management [71].

7.7 Social search engines

A social search engine is a type of search engine that generates the answer to user queries and evaluates its relevance based on the interactions or contributions of other users [22]. Before the social search engine will be able to provide this type of answer, the appropriate information about user preferences must be gathered. This can be done for example by social bookmarking, see Section 7.8, or the system can ask the user whether the answer to the query is relevant or not. Every user of this kind of service must possess their own account (the internet identity) that enables the personalization to be permanent. People can exchange the information about their preferences so that we can create a social network connecting people with similar interests. Many forms of social search may be distinguished, from the simple shared bookmarking or tagging of the content to more sophisticated approaches that combine human intelligence with computer Information Retrieval algorithms.

In contrary to machine–based searching, e.g. using Google's PageRank, the social approach gives the opportunity for more personalized and in consequence probably more relevant answers to queries asked by the specific individuals [54].

There are some start–up portals for social searching like *Wikia Search*, *Mahalo.com*.

7.8 Social bookmarking and cataloguing

Social bookmarking enables users to store, organize, search and last but not least share with other users bookmarks of web pages [65]. Some popular sites serving as social bookmarking are: *del.icio.us* [3, 49], *Furl*, *Google Bookmarks*, *Diigo*.

The bookmarks depending on the features of a given service can be saved privately, shared only with some chosen individuals, groups or only inside the certain network. Most social bookmarking services enable users to organize their bookmarks with the shared tags and/or folders. They also enable viewing bookmarks associated with the given tag. Most of the bookmarking services provide also additional features such as the possibility of rating and commenting on bookmarks, the ability to import and export bookmarks from web browsers, emailing of bookmarks, web annotation, and building groups, etc.

All above features enable to extract social networks within social bookmarking sites, in which user logins are the internet identities and separate relationships can be derived from different shared meeting objects like bookmarks, tags, folders, groups, etc., see Section 5.3.

Social cataloguing is a concept similar to social bookmarking. Its main aim is to provide users the opportunity to catalogue things they possess, e.g. books, music, films,

etc. Each user creates and shares with other members the description of items they want to catalogue.

Some popular sites serving as social cataloguing are: (i) for books – *LibraryThing*, *Shelfari*, *Goodreads*, *Anobii*, *Books iRead*, (ii) for music – *Discogs*, *Rate Your Music*, *Last.fm*, (iii) for movies – *Flixster*; (iv) for scholary citations – *Bibster*, *CiteULike*, *Connotea*.

Users share the metadata about the items as well as interact and cooperate with each other by improving their descriptions. The social network can be created based on the description of the item that can be treated as an object that connects users who participate in its creation and maintenance.

An example of social bookmarking and simultaneously social cataloguing service is *CiteULike*, which facilitates sharing scientific references among researchers. It also supports import of bibliographical descriptions directly from some most common sites such as *Amazon.com*, *SpringerLink* or *ScienceDirect*.

7.9 Homepages

A single homepage is the web site that contains and provides information about a specific person. Homepages are usually maintained by users to whom these pages belong; they can add and update information about their life, work and interests. Users can also incorporate some hyperlinks to others' homepages into their HTML contents. These external, linked homepages can belong to their friends, family members, partners they cooperate with or even other people being considered as interesting by the creator. The URL address of the homepage can be treated as the internet identity of the person this web site belongs to. Moreover, all links to others homepages are signs of direct relationships from the given internet identity to all others it links to. Hence, the system of homepages is an example of the homogenous social network (*HSN*). Note that the relationships are asymmetrical; it means that the target homepages do not have to contain the reverse links. Besides, in such a network it is not possible to assess the strength of relations, so there are only two states: a relation either exists or not.

7.10 Knowledge sharing systems

Knowledge markets (*Experts-Exchange*, *Mahalo Answers*, *Yahoo! Answers* [49], *Knowledge Search*, *ChaCha.com*, *Answerly.com*) are examples of social networks that enable users, on the one hand, to post a request and set a virtual price for the relevant answer as well as, on the other hand, to answer the questions that others have asked and get reward for the correct answers. Moreover, users have the opportunity to rate and comment the answers they have received from others. Based on each type of these activities the separate layer in system–based social network can be created.

The knowledge in such systems is treated as regular, tangible goods. The currency that is used to pay for the most relevant answers are points as in *Experts-Exchange* or virtual currency as in the case of *Mahalo Answers* where the binding currency is Mahalo Dollar. Nevertheless, none of the enumerated systems enables to pay using real money for the valid information and the only award for the correct answers is the high prestige among other network members. *ChaCha* and *Answerly* are the examples of the systems, in which the experts are paid for their answers but people can still use the system for free. Yet another system was proposed in [44], in which social relationships are used to gather the latent knowledge from experts within an organization.

7.11 Virtual worlds and multiplayer online games

The virtual worlds and multiplayer online games (*Second Life*, *Sims*, *World of Warcraft*) are the examples of systems, in which users maintain their own avatars that represent them in the virtual world. People can create not only their avatar but also the whole neighbourhood they want to live in. This leads to the situation that social entities create for themselves the second life that is parallel to the real one. In online games, users can cooperate with other players by attending common missions [21]. Sometimes, there is even a situation that a task cannot be accomplished by a single person. Thus, it is inevitable that users merge into groups.

These systems somehow map the real world to the virtual one. That means that the virtual world social networks can correspond to the real world social networks. The internet identities will be the avatars and any cooperation between them is the basis to create the relationship between the network members.

7.12 Collaborative authoring systems – wikis

Wikis are yet another example of social networks on the Internet where users represented by their internet identities collaborate in order to create the common content. For example, in *Wikipedia* one user initiates the work on the specific term and other users can contribute by changing and improving the term description. Such cooperation, while creating the content, provides the opportunity to obtain the outcome, which quality is higher than in the case of a single author. An article or a term description in collaborative authoring systems plays the similar role to the one that plays description of an item in social cataloguing systems, i.e. it is a meeting object that connects people who are involved in the process of creating it [42].

7.13 Friend of a friend project

The general purpose of *Friend Of A Friend* (*FOAF*) project is quite simple: to built such representation of users, their activities, and acquaintances that can be processed by a computer. In order to achieve this goal the appropriate *FOAF* machine–readable ontology was developed. These comprehensive users' profiles, which also include the links to their friends, create a homogenous social network *HSN* [8]. From the technical point of view, *FOAF* files are defined using Web Ontology Language (OWL) being an extension of Resource Description Framework (RDF). The usage to *FOAF* is free so every internet user can exploit it to create their personal profile and to define the relationships maintained by this person. People are mostly using the *FOAF* format to put their personal data into an RDF file and to publish it on their homepages. Next, web crawlers gather and aggregate the information, for example *SECO* [29]. Moreover, every participant possesses a unique identity – *OpenID* that is used while processing the relationships defined by this user. This enables computers to find people who are similar to each other or who maintain similar relationships. Recently, many social networking sites SNS have started to support the *FOAF* format to exchange user profile information [26].

7.14 Complex communications systems

The popularity and diversity of the instant messengers was the inspiration to create some integrated services such as *eBuddy* or *Miranda* that enable to join together separate user accounts from different communication systems. For example, *eBuddy*, which is web and mobile messenger, supports multiple instant messaging services such as *Windows Live*

Messenger, *Yahoo!*, *AIM*, *Google Talk* (*GTalk*), *Facebook* and *MySpace IM*, *ICQ* within one interface. The *eBuddy* system utilizes its own *eBuddy ID* (a joint internet identity) to authenticate its users.

7.15 Comparison

The enumerated and described above categories of social networks on the Internet can be compared in many separate aspects (see Tables 3 and 4). The layers within ties, see Section 5.4, which can be distinguished within each of the network classes are presented in Table 3. These layers are derived from different types of user activities within the given category of social networks, including direct mutual communication between users via different communication channels, similar and shared activities towards a certain meeting object (e.g. common usage of tags, commenting or adding to favourites of the same objects, being a part of the same group, etc), contact lists, or even similarities between users' profiles they maintain.

Based on the analyses of the characterized categories of social networks on the Internet their key features were identified and the comparison of these characteristics between different social network classes is presented in Table 4.

8 Discussion

The main rationale that stays behind this survey is lack of the coherent and comprehensive approach to social networks on the Internet from research point of view. This is a relatively new concept and although a lot of studies have been conducted in this area, no work has been done to grasp the whole idea. Usually, researchers focus on one specific example of social network, e.g. email or blogs and do not present them in wider context. This is understandable, as most of the work is focused on one specific application area. However, authors of this paper claim that general overview of social networks on the Internet is needed since it gives others a wider horizon of their work and help to systemize research in this field. This will also enable new researchers in this area to find in one place complete information about this topic together with wide range of literature review.

The whole research presented in the paper was divided into four main topics:

- Definition of social networks on the Internet with different level of complexity.
- Detailed description of the internet identity and internet relationship concepts together with virtual identity and tie concepts, which are natural extension of two previous ones respectively.
- Comprehensive and coherent taxonomy of social networks on the Internet
- Identification and profiles of various categories of social networks that can be discovered within different services available on the Internet

Some parts of each of the enumerated above topics are addressed in the literature, however, it can be noticed that very often the proposed definitions and taxonomies are not general enough to cover the whole diversity of the recent Internet.

In the case of the social network on the Internet, the definitions of different researchers rather refer to the intuition that stays behind this concept, i.e. group of people who interact with each other using various services on the Internet, than to the formal definition. This ambiguity of the social networks on the Internet concept and lack of the systematic approach

Table 3 Layers in system–based social networks on the Internet.

Category of the service/system	Layers in the social network (directness of relationships, see section 5.3)
Email service	a) Communication: sent/received emails (direct)
	b) Address books (direct)
Instant messenger	a) Communication: a separate layer for each communication channel, e.g. text messages, VoIP, SMS, video conferences, etc. (direct)
	b) Address books (direct)
	c) Profile-based similarities (indirect)
Blog service	a) References to other blogs (blogrolls)[a] (direct)
	b) Comments, a separate layer for commentators and commentator–author (quasi–direct, meeting object: a post)
	c) Common usage of tags/keywords/categories (quasi–direct, meeting object: a tag, keyword, category)
	d) Profile-based similarities (indirect)
Social networking site	a) Communication: a separate layer for each communication channel, e.g. sent/received emails, video conference, etc. (direct)
	b) Contact lists (direct)
	c) Groups of interest, school classes (quasi–direct, meeting object: a group)
	d) Profile–based similarities (indirect)
Multimedia sharing system	a) Contact lists (direct)
	b) Comments and favourites: a separate layer for commentators and commentator-author of the shared objects as well as for favourite–favourite and favourite–author (quasi–direct, meeting object: a multimedia object)
	c) Common usage of tags (quasi–direct, meeting object: a tag)
	d) Common groups (quasi–direct, meeting object: a group)
Auction system	a) Auction: a separate layer for seller-bidder, seller-buyer/commentator and bidder-bidder (quasi–direct, meeting object: an auction, a bid, transaction, or comment to the auction)
	b) Invitation to the restricted auction: inviting-invited, invited-invited (quasi–direct, meeting object: an auction)
	c) Communication: questions and answers (emails) referring the auction (direct)
	d) Removal of unwanted bids by the seller, a negative relationship (direct)
	e) Profile–based similarities (indirect)
Social search engine	a) Profile–based similarities (indirect)
Social bookmarking and cataloguing	a) Shared bookmarks or item descriptions (quasi–direct, meeting object: bookmark, bibliographical description)
	b) Profile–based similarities (indirect)
Homepages	a) References (hyperlinks) to other homepages (direct)
Knowledge Sharing Systems	a) Questioning: query asking person – answering expert (quasi–direct, meeting object: a question)
	b) Commenting and rating: commentator-expert (quasi–direct, meeting object: an answer)
Virtual Worlds & online games	a) Shared activities/collaboration: participant–participant (quasi–direct with either the same or different role, meeting object: a shared activity, e.g. common mission)
	b) Shared objects: user–user (quasi–direct, usually with the same role, meeting object: a shared object, e.g. a weapon)
	c) Teams and groups: member–member, leader-member (quasi–direct, meeting object: a common group/team)

Table 3 (continued).

Category of the service/system	Layers in the social network (directness of relationships, see section 5.3)
	d) Communication: a separate layer for each message type, e.g. sent/received orders, invitations, comments, etc. (direct)
Collaborative Authoring Systems – Wikis	a) Collaborative working on common topics/contents: co-author – co-author (quasi–direct, meeting object: a topic)
	b) Discussion: (quasi–direct, meeting object: a threat of the discussion)
FOAF	a) References to profiles of other FOAF users (direct)
	b) Profile–based similarities (indirect)

Blogrolls can be seen as the address books in case of the email service or instant messengers

to classify these networks resulted in three definitions proposed in this paper, the definition of (i) a homogenous, (ii) a system-based and (iii) an internet multisystem social network, see Section 3.1. The internet multisystem social network is a very important contribution of this paper as it can encapsulate not only many different relationships but within such a network, a single user can maintain more than one representation called internet identity, see Sections 3.1 and 5.4. These three definitions together constitute a framework that enables to define and classify each of the social network extracted from the data available in the systems presented in Section 7.

This can also be a reference for future explorations, in which researchers analyzing one type of network can compare it with other networks of the same type, e.g. homogenous social network with relations of a given type. It appears that upcoming studies on social network will try to cover more and more complex environments created by the increasing interoperability of the internet-based services. Authors believe that multilayered, system-based social networks as well as internet multisystem social networks may be very interesting subjects for new investigations.

9 Conclusions

The coherent analysis and classification of social networks on the Internet have been provided in this survey. The number of internet services, in which users get into various relationships continually increases, thus the described categories of social networks do not exhaust the domain.

Starting from the analysis of social networks on the Internet, many new potential opportunities arise. They enable to discover and utilize specific knowledge lying dormant within these networks. On the other hand, a couple of new problems come up during such analyses. The main issues that should be addressed are data security and reliability. It includes identity theft or privacy protection [7, 59] as well as spammers or fake identities. Another issue while analysing large amounts of data is the efficiency of processing. The problems that are hard to overcome are the integrations of social networks derived from separate systems, identification the group identities as well as merging of individual internet identities into combined virtual identities.

The knowledge that is contained in the data about users and their relationships belonging to social networks on the Internet can be utilized for marketing purposes, in recommender

Table 4 Features of the system–based social networks.

No.	Features	Email service	Instant messenger	Blog service	Social networking site	Multimedia sharing system	Auction system	Social search engine	Social bookmarking and cataloguing	Homepages	Knowledge Sharing Systems	Virtual Worlds & online games	Collaborative Authoring Systems – Wikis
1.	Public availability of data about social network	No	No	Yes	Medium[a]	Yes	Medium[b]	No	Yes	Yes	Yes	No	Yes
2.	The build-in function for searching for members to get into new relationships with	No	Yes	Yes	Yes	Yes	Yes	No	No	No	No	Yes/No	No
3.	Visibility of own incoming relationships within the system GUI[c]	No	No	No	Yes	No	No	No	No	No	No	No	No
4.	Visibility of own outgoing relationships within the system GUI[d]	Partly[e]	Partly[f]	Partly[g]	Yes	Partly[h]	Yes	No	No	Yes	No	Yes/No	No
5.	The possibility of defining the type of relationship	No	No	No	Yes	No	No	N/A	N/A	No	No	No	No
6.	The awareness level of being in the incoming relation (High/Medium/None)	High/None[i]	High/None[j]	Medium	High	Medium	High	None	None	None	Medium	High	Medium
7.	The awareness level of being in the outgoing relation (High/Medium/None)	High	High	High	High	High	High	None	None	High	Medium	High	Medium
8.	Does the user have to directly specify that is in the relationship with someone?	No	No	Yes/No	Yes	Yes/No	Yes/No	No	No	Yes	No	No	No
9.	Does the user have to directly state that breaks up the relationship with someone?	No	No	No	Yes	No	No	No	No	Yes	No	No	No
10.	Dedicated tools for relationships maintenance	Yes/No[k]	Yes/No[11k]	No	Yes	Yes/No[l]	No	No	No	No	No	Yes/No	No
11.	Can time factor for relationships be analyzed? (Yes/ Sometimes/No)	Sometimes [m]	Sometimes [13m]	Sometimes[n]	Yes	Sometimes	Yes	No	No	No	Yes	Yes	Yes
12.	Do groups of *IID*s exist? [o]	Yes	Yes	Yes	No	No	No	No	No	Yes	No	No	No

Table 4 (continued).

No.	Features	Email service	Instant messenger	Blog service	Social networking site	Multimedia sharing system	Auction system	Social search engine	Social bookmarking and cataloguing	Homepages	Knowledge Sharing Systems	Virtual Worlds & online games	Collaborative Authoring Systems – Wikis
13.	Is the data about relations centrally maintained?	Yes[p]	Yes/No[q]	Yes[r]	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes

[a] The access to the data about the whole network is usually restricted, i.e. there is only access to the relationships of the recently searched people

[b] Data about some user activities are available only for the system operator, e.g. removal of unwanted bids, communication between users, some historical auctions

[c] It means that there exists a function in the system that directly displays the list of the connections from other users to the given one

[d] It means that there exists a function in the system that directly displays the list of the connections from the given user to others

[e] Relationships derived from address books are visible whereas those based on outgoing emails are not

[f] Relationships derived from contact lists are visible whereas those created upon communication only are not

[g] In case of blogrolls (direct reference layer) – yes, in case of comments – no

[h] Relationships derived from contact lists are visible whereas those created upon communication only are not

[i] Users are aware of emails they receive but the content of address books maintained by others is usually unknown

[j] Users are aware of incoming communication but the content of address books maintained by others is usually unknown

[k] "Yes" refers only address books

[l] "Yes" refers only contact lists

[m] Communication data is time-stamped while address books are not

[n] In case of favourite list there is no information about when the object was added to favourites whereas the time when the multimedia (meeting) object was commented by the given person is usually provided

[o] There is possibility to use a single account by a group of people in every system, however, it may be uncommon for some services

[p] Yes, but the data can be centrally maintained only within one email server/domain, e.g. *@gmail.com*

[q] It depends on the provider of the service, e.g. in case of the Skype system, the data about contact lists are centrally maintained whereas the data about communication are not stored

[r] Yes, but only within a single blog service, e.g. *WordPress*

systems, in trust management processes, to support collaborative work, and even to identify terrorist networks, etc.

Acknowledgments This work was supported by The Polish Ministry of Science and Higher Education, the development project, 2009-11, and research project, 2010-13.

References

1. Adamic, L.A., Adar, E.: Friends and neighbors on the web. Soc. Network **25**(3), 211–230 (2003)
2. Amaral, L.A.N., Scala, A., Barthelemy, M., Stanley, H.E.: Classes of small–world networks. Proc. Natl. Acad. Sci. USA **97**, 11149–11152 (2000)
3. Bao, S., Yang, B., Fei, B., Xu, S., Su, Z., Yu, Y.: Social propagation: boosting social annotations for web mining. World Wide Web **12**(4), 399–420 (2009)
4. Barnes, J.A.: Class and committees in a Norwegian Island Parish. Hum. Relat. **7**, 39–58 (1954)
5. Beyene, Y., Faloutsos, M., Chau, D.H., Faloutsos, C.: The eBay graph: How do online auction users interact? In proceedings of INFOCOM Workshops, IEEE Computer Society, 1–6 (2008)
6. Boyd, D.M.: Friendster and Publicly Articulated Social Networking, In Proceedings of CHI '04: CHI '04 extended abstracts on Human factors in computing systems, pp. 1279–1282. ACM Press (2004)
7. Boyd, D.M., Ellison, N.B.: Social network sites: definition, history, and scholarship. J. Comput-Mediat Comm. **13**(1), 210–230 (2007)
8. Brickley, D. Miller, L.: FOAF Vocabulary Specification 0.91. Namespace Document 2 November 2007 - OpenID Edition (2007), http://xmlns.com/foaf/spec/, Accessed 10.11.2009
9. Carrington, P., Scott, J., Wasserman, S.: Models and Methods in Social Network Analysis. Cambridge University Press, Cambridge (2005)
10. Castells, M.: Virtual Communities or Network Society? In The Internet galaxy: reflections on the Internet, business, and society, pp. 116–136. Oxford University Press (2001)
11. Cattell, V.: Poor people, poor places, and poor health: the mediating role of social networks and social capital. Soc. Sci. Med. **52**(10), 1501–1516 (2001)
12. Caverlee, J., Webb, S.: A Large-Scale Study of MySpace: Observations and Implications for Online Social Networks, In Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media, pp. 36–44. AAAI Press (2008)
13. Chiu, P.Y., Cheung, C. M. K. Lee, M. K. O.: Online Social Networks: Why Do "We" Use Facebook?, In Proceedings of the First World Summit on the Knowledge Society, Communications in Computer and Information Science 19, Springer, 67–74, (2008)
14. Creese, J., Cribb, J., Spicer, J.: Social Networking: never mind the students, what about us? Use of Social Networking Softwares for professional networking and development for library staff, In Proceedings of Beyond The Hype 2008: Web 2.0 Symposium (2008), http://espace.library.uq.edu.au/eserv/UQ:124191/social_network_library_profession.pdf, Accessed 31 August 2009
15. Culotta, A., Bekkerman, R., McCallum, A.: Extracting social networks and contact information from email and the Web, In Proceedings of the First Conference on Email and Anti-Spam (2004)
16. Degenne, A., Forse, M.: Introducing Social Networks. SAGE Publications Ltd, London (1999)
17. Desautels, B.: Social networking: LinkedIn.com. J Leader. Stud. **2**(2), 103–104 (2008)
18. Dimicco, J.M., Millen, D.R.: Identity Management: Multiple Presentations of Self in Facebook, In Proceedings of the 2007 International ACM SIGGROUP Conference on Supporting Group Work, pp. 383–386. ACM Press (2007)
19. Dimicco, J., Millen, D.R., Geyer, W., Dugan, C., Brownholtz, B., Muller M.: Motivations for Social Networking at Work, In Proceedings of the Computer Supported Cooperative Work 2008 Conference, ACM Digital Library, pp. 711–720. (2008)
20. Donath, J.S.: Identity and deception in the virtual community, Chapter 2. In: Smith, M.A., Kollock, P. (eds.) Communities in Cyberspace, pp. 29–59. Routledge, London & New York (1999)
21. Ducheneaut, N., Yee, N., Nickell, E., Moore, R.J.: "Alone together?": exploring the social dynamics of massively multiplayer online games, In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, pp. 407–416. (2006)

22. Evans, B.M., Kairam, S., Pirolli, P.: Do your friends make you smarter?: an analysis of social strategies in online information seeking. Inform. Process. Manag. **46**(6), 679–692 (2010)
23. Flake, G., Lawrence, S., Lee Giles, C.: Efficient identification of web communities, In Proceedings of the Sixth ACM Conference on Knowledge Discovery and Data Mining, pp. 150–160. (2000)
24. Golbeck, J., Hendler, J.: FilmTrust: movie recommendations using trust in web-based social networks, In Proceedings of Consumer Communications and Networking Conference, IEEE Conference Proceedings 1, 282–286 (2006)
25. Golbeck, J.: Computing and Applying Trust in Web-Based Social Networks, Dissertation Submitted to the Faculty of the Graduate School of th Universtity of Maryland, College Park in partial fulfilment of the requirements for the degree of Doctor of Philosophy (2005)
26. Golbeck, J., Rothstein, M.: Linking Social Networks on the Web with FOAF: A Semantic Web Case Study, In Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, pp. 1138–1143. AAAI Press (2008)
27. Gross, R., Acquisti, A.: Information revelation and privacy in online social networks, In Proceedings of 2005 ACM workshop on Privacy in the electronic society, pp. 71–80. ACM Press (2005)
28. Han, L., Cen, S., Ma, J., Ma, Y.: Identify intimate social relationship in blogsphere, lecture notes of the institute for computer sciences. Soc. Inform. Telecommun. Eng. **18**, 14–25 (2009)
29. Harth, A.: SECO: mediation services for semantic web data, IEEE intelligent systems. Spec. Issue Semant Web Chall. **19**(3), 66–71 (2004)
30. Hill, R., Dunbar, R.: Social network size in humans. Hum. Nat. **14**(1), 53–72 (2002)
31. Howard, B.: Analyzing online social networks. Commun. ACM **51**(11), 14–16 (2008)
32. Huberman, B.A., Romero, D.M., Wu, F.: Social networks that matter: Twitter under the microscope. First Monday **14**(1) (2009)
33. Jung, J.J., Juszczyszyn, K., Nguyen, N.T.: Centrality measurement on semantically multiplex social networks: divide-and-conquer approach. Int. J. Intell. Inform. Database Syst. **1**(3/4), 277–292 (2007)
34. Kazienko, P., Bródka, P., Musial, K., Gaworecki, J.: Multi-layered Social Network Creation Based on Bibliographic Data. SocialCom-10, The Second IEEE International Conference on Social Computing, SIN-10 Symposium on Social Intelligence and Networking, IEEE Computer Society Press, 407–412 (2010)
35. Kazienko, P., Musiał, K., Kajdanowicz, T.: Multidimensional social network and its application to the social recommender system. IEEE Trans. Syst. Man Cybern. A Syst. Hum. **41**(4), 746–759 (2011)
36. Kazienko, P., Musiał, K., Zgrzywa, A.: Evaluation of node position based on email communication. Contr. Cybern. **38**(1), 67–86 (2009)
37. Kazienko, P.: Expansion of Telecommunication Social Networks, In Proceedings of the fourth International Conference on Cooperative Design, Visualization and Engineering, Springer Verlag, Lecture Notes in Computer Science **4674**, 404–412 (2007)
38. Kennedy, H.: Beyond anonymity, or future directions for internet identity research. New Media Soc. **8**(6), 859–876 (2006)
39. Kim, W., Jeong, O.-R., Lee, S.-W.: On social web sites. Inform. Syst. **35**(2), 215–236 (2010)
40. King, I.: Introduction to social computing, database systems for advanced applications. Lect. Notes Comput. Sci. **5982**, 482–484 (2010)
41. Kinsella, S., Breslin, J.G., Passant, A., Decker, S.: Applications of Semantic Web Methodologies and Techniques to Social Networks and Social Websites, Reasoning Web, In Proceedings of the 4th International Summer School 2008, Tutorial Lectures, Lecture Notes in Computer Science 5224, Springer, 171–199 (2008)
42. Klamma, R.: Wikis as Social Networks: Evolution and Dynamics, In The 2nd SNA-KDD Workshop on Social Network Mining and Analysis, ACM (2008)
43. Krishnamurthy, B., Wills, C.E.: On the leakage of personally identifiable information via online social networks. ACM SIGCOMM Comput. Comm. Rev. **40**(1), 112–117 (2010)
44. Kukla, G., Kazienko, P., Bródka, P., Filipowski, T.: Recommendation Boosted Query Propagation in the Social Network, In Proceedings of The Second International Conference on Social Informatics, SocInfo'2010, Lecture Notes in Artificial Intelligence 6430, Springer, 113–124 (2010)
45. Kumar, R., Novak, J., Raghavan, P., Tomkins, A.: On the bursty evolution of blogspace. World Wide Web **8**(2), 159–178 (2005)
46. Kwak, H., Lee, C., Park, H., Moon, S.B.: What is Twitter, a social network or a news media? in Proceedings of the 19th International Conference on World Wide Web, WWW 2010, ACM, 591–600 (2010)
47. Lazega, E.: The Collegial Phenomenon. The Social Mechanism Of Co–Operation Among Peers In A Corporate Law Partnership. Oxford University Press, Oxford (2001)
48. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Life in the network: the coming age of computational social science. Science **323**(5915), 721–723 (2009)

49. Leskovec, J., Backstrom, L., Kumar, R., Tomkins, A.: Microscopic evolution of social networks, In Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 462–470 (2008)
50. Lin, C.-L., Tang, H.-L., Kao, H.-Y.: Utilizing social relationships for blog popularity mining. Lect. Notes Comput. Sci. **5839**, 409–419 (2009)
51. Marx, K.: Selected Writings in Sociology & Social Philosophy, tr. T.B. Bottomore, New York, McGraw–Hill ([1857] 1956)
52. Milgram, S.: The small-world problem. Psychol. Today **2**, 60–67 (1967)
53. Millen, D., Feinberg, J., Kerr, B.: Social bookmarking in the enterprise, ACM Press, Queue 3, 9, 28–35 (2005)
54. Mizzaro, S., Vassena, L.: A social approach to context-aware retrieval. World Wide Web **14**(4), 377–405 (2011)
55. Montgomery, J.: Social networks and labor-market outcomes: toward an economic analysis. Am. Econ. Rev. **81**(5), 1407–1418 (1991)
56. Morris, M.: Sexual network and HIV. AIDS **11**, 209–216 (1997)
57. Narayanan, A., Shmatikov, V.: De-anonymizing Social Networks, In proceedings of the 30th IEEE Symposium on Security and Privacy, IEEE Computer Society, 173–187 (2009)
58. Newman, M.E.J.: The structure of scientific collaboration networks, In Proceedings of the National Academy of Sciences of the United States of America, 98, 404–409 (2001)
59. O'Murchu, I., Breslin, J.G., Decker, S.: Online Social and Business Networking Communities, In Proceedings of the Workshop on Application of Semantic Web Technologies to Web Communities 2004, CEUR Workshop Proceedings, (2004), 107, http://ftp.informatik.rwth-aachen.de/Publications/CEUR-WS/Vol-107/paper2.pdf, Accessed 10 November 2009
60. Pagel, M., Erdly, W., Becker, J.: Social networks: we get by with (and in spite of) a little help from our friends. J. Pers. Soc. Psychol. **53**(4), 793–804 (1987)
61. Parameswaran, M., Whinston, A.B.: Research issues in social computing. J. Assoc. Inform. Syst. **8**(6), 336–350 (2007)
62. Rak, J.: The digital queer: weblogs and internet identity. Biography **28**(1), 166–182 (2005)
63. Recordon, D., Reed, D.: OpenID 2.0: a platform for user-centric identity management, In Proceedings of the 2006 Workshop on Digital Identity Management, ACM, 11–16 (2006)
64. Robins, G.L., Alexander, M.: Small Worlds Among Interlocking Directors: Network Structure and Distance in Bipartite Graphs, Computational & Mathematical Organization Theory 10, 1, Kluwer Academic Publisher, 69–94 (2004)
65. Scott, G., Huberman, B.A.: Usage Patterns of Collaborative Tagging Systems, Journal of Information Science 32, 2, 198–208 (2006), http://www.hpl.hp.com/research/idl/papers/tags/index.html, Accessed 10 November 2009
66. Swamynathan, G., Wilson, C., Boe, B., Almeroth, K., Zhao, B.Y.: Do social networks improve e-commerce?: a study on social marketplaces, In Proceedings of the First Workshop on Online Social Networks, ACM, 1–6 (2008)
67. Travers, J., Milgram, S.: An experimental study of the small world problem. Sociometry **32**(4), 425–443 (1969)
68. Trusov, M., Bodapati, A.V., Bucklin, R.E.: Determining influential users in internet social networks. J. Market. Res. **47**(4), 643–658 (2010)
69. Walker, K.: It's difficult to hide it: the presentation of self on internet home pages. Qual. Sociol. **23**(1), 99–120 (2000)
70. Wang, F.Y., Carley, K.M., Zeng, D., Mao, W.: Social computing: from social informatics to social intelligence. IEEE Intell. Syst. **22**(2), 79–83 (2007)
71. Wang, J.C., Chiu, C.C.: Recommending trusted online auction sellers using social network analysis. Expert Syst. Appl. **34**(3), 1666–1679 (2008)
72. Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, New York (1994)
73. Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., Haythornthwaite, C.: Computer networks as social networks: collaborative work, telework, and virtual community. Annu. Rev. Sociol. **22**(1), 213–238 (1996)
74. Wood, A.F., Smith, M.J.: Online Communication: Linking Technology, Identity & Culture, Secondth edn. Routledge, New York (2004)
75. Yang, W.S., Dia, J.B., Cheng, H.CH., Lin, H.T.: Mining Social Networks for Targeted Advertising, In Proceedings of the 39th Hawaii International International Conference on Systems Science, 6, IEEE Computer Society, 137a (2006)
76. Yeung, C.-m.A., Liccardi, I., Lu, K., Seneviratne O., Berners-Lee T.: Decentralization: The Future of Online Social Networking, In Proceedings of W3C Workshop on the Future of Social Networking, http://www.w3.org/2008/09/msnws/papers/decentralization.pdf (2009)