Properties of Bridge Nodes in Social Networks

Katarzyna Musiał and Krzysztof Juszczyszyn

Wroclaw University of Technology, Institute of Computer Science,
Wybrzeze Wyspianskiego 27, Wroclaw, Poland
{katarzyna.musial,krzysztof.juszczyszyn}@pwr.wroc.pl

Abstract. The main goal of the paper is to describe the properties of the nodes within a social network that connect the peripheral nodes and peripheral groups with the rest of the network. These nodes are usually called bridging nodes or simply bridges. All the experiments are carried out on the real data, so-called Thurman network. First, the regular cliques, peripheral nodes and peripheral cliques from a network are extracted and then the bridging nodes identified. Afterwards for all nodes their characteristic features, such as social position and degree of nodes, are calculated. Finally, we try to find the correlation between nodes centrality and their degree and the fact if given node is a bridge or not.

Keywords: bridging node, regular clique, peripheral node, peripheral clique, social position.

1 Introduction

The analysis of the characteristics of the nodes that connect the whole network with the peripheral nodes or peripheral cliques that are loosely connected with the rest of the network is a very interesting research problem. The issue of identifying bridges within a given network is a complex and resource consuming task because it involves an extensive analysis of the groups and cliques existing within a given network [12], [10], [4]. Bridges can be seen as the nodes without which the network will split into two or more subgroups. This concept is similar to the idea of weak ties [3], [6], [1] that also tend to be vital bridges between the two densely knit clumps. In this work bridges are defined as the nodes that (i) connect regular cliques with the peripheral nodes or (ii) connect regular cliques with the peripheral cliques that are loosely connected with the rest of the network. The explanations of such concepts as regular clique, peripheral clique, bridging node and peripheral node as well as the whole method for bridges properties analysis is presented. By defining the properties of the bridges we will be able to identify them without the complex calculations.

2 Method for Bridges' Properties Analysis

The main goal of the paper is to identify bridges within the social network and describe their characteristic features. The whole process is presented in the

N.T. Nguyen, R. Kowalczyk, and S.-M. Chen (Eds.): ICCCI 2009, LNAI 5796, pp. 357–364, 2009.

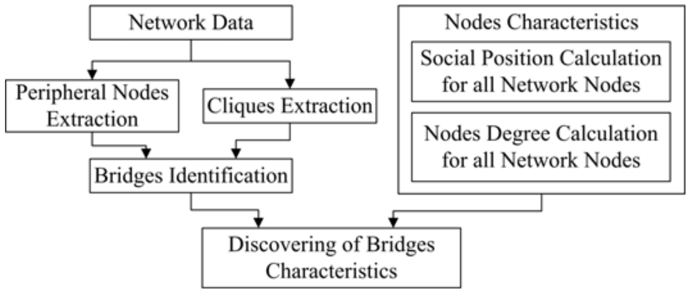

Fig. 1. Crucial steps of the proposed method of bridges properties analysis

Figure 1. The method is split into two main groups of activities: (i) the process of bridges identification and (ii) the process of the network nodes properties calculation. Note that the network is represented as a weighted and directed graph.

In the former process, the first step is to extract all cliques existing in a given network. The clique is defined as "a maximal complete subgraph of three or more nodes" [12]. It means that a clique consists of a subset of nodes, all of which are connected to all nodes of the clique. Additionally, in the rest of the network does not exist even one node that is connected to all nodes that belong to the analyzed clique [9], [5]. In the same time, the peripheral nodes and peripheral cliques are identified. The former ones are nodes that do not belong to any clique that exist within an analyzed network. The latter ones are cliques that are loosely connected with the rest of the network, i.e. they do not have any common node with the rest of the extracted cliques. The set of cliques that does not include the peripheral cliques is called a set of regular cliques. Afterwards the bridging nodes can be discovered. Bridges are nodes that belong to the regular clique and connect it with peripheral node or peripheral clique.

The second set of activities is connected with analyzing the characteristic features of the network nodes. Different characteristics from the node perspective can be analyzed, e.g. centrality, prestige, density, social capital [12], [4]. In the presented research, we take into consideration two of them: centrality index named social position [7], [8] as well as the number of incoming, outgoing and mutual relationships of the node [12], [2]. Social position function $SP(x)$ of member x in the social network, respects both the value of node positions of all other network members as well as the level of their activities in relation to x [7], [8]:

$$SP(x) = (1 - \varepsilon) + \varepsilon \cdot \sum_{i=1}^{m_x} (SP(y_i) \cdot C(y_i \rightarrow x)) \tag{1}$$

where: ε – the constant coefficient from the range $[0, 1]$; y_i — xs acquaintances, i.e. the members who are in direct relationship to x: $C(y_i \rightarrow x) > 0$; $C(y_1 \rightarrow$

x),...,$C(y_m \to x)$ – the commitment function that denotes the contribution in activity of y_1,...,y_m directed to x; m_x — the number of x's acquaintances.

The importance of the node in the weighted and directed network, expressed by the social position function, tightly depends on the strength of the relationships that other members of the network maintain with the given node as well as on the social positions of these members. In other words, the member's social position is inherited from others but the level of inheritance depends on the activity of the members directed to this person, i.e. intensity of common interaction, cooperation or communication.

The last stage of the method is to find the features of the bridging nodes that are their specific and individual properties and distinguish them from the other network nodes. If it is possible to find such features, then it will be no need to perform the extensive calculations in order to locate the bridging nodes in the network structure.

3 Experiments

The experiments were conducted on the Thurman office social network that is a non–symmetrical network of 15 people who worked in one company. Thurman spent 16 months observing the interactions among employees in the office of a large corporation [11].

The adjacency matrix for the Thurman network is presented in Table 1 where values 1 represent the existence of the connection between two users. In order to obtain the values of commitment function (that are needed to calculate the social position values, see Equation 1) for each individual, value one — from the

Table 1. The values of commitment function within the Thurman network

Member	12	1	5	2	8	6	3	10	4	9	7	11	13	14	15
12.Emma	–	–	1	–	1	1	–	1	–	1	–	1	1	1	1
1.Ann	–	–	1	1	1	1	1	–	–	1	–	1	1	–	–
5.Pete	1	1	–	1	1	1	1	1	1	1	1	1	1	1	1
2.Amy	–	1	1	–	1	1	1	–	1	–	–	–	–	–	–
8.Lisa	1	1	1	1	–	1	1	–	–	1	–	–	–	–	–
6.Tina	–	1	1	1	1	–	1	–	–	–	–	–	–	–	–
3.Katy	–	1	1	1	1	1	–	–	–	–	–	–	–	–	–
10.Minna	1	–	–	1	1	–	–	–	1	–	1	–	–	–	–
4.Bill	–	–	–	1	–	–	–	1	–	–	1	–	–	–	–
9.President	1	1	1	1	1	1	1	1	1	–	1	1	1	1	1
7.Andy	–	–	1	–	–	–	–	1	1	–	–	–	–	–	–
11.Mary	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–
13.Rose	1	1	–	–	–	–	–	–	–	–	–	–	–	–	–
14.Mike	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
15. Peg	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–

original matrix Table 1 — is divided by the number of members relationships, e.g. Emma communicates with nine members so her contribution of activity to each of her acquaintances equals $\frac{1}{9}$.

3.1 Process of Bridges Extraction

First step of the method is to extract the cliques existing within the network. Note, that the relationships within the Thurman network are weighted and directed. It means that group can be called clique if and only if all the relationships between members are mutual. Four cliques were extracted from the Thurman network (the numbers indicate the ID of the specific user, see Table 1):

1. $C_1 = \{Ann, Amy, Lisa, Katy, Tina, Pete\} = \{1, 2, 8, 3, 6, 5\}$
2. $C_2 = \{Ann, Lisa, Pete, President\} = \{1, 8, 5, 9\}$
3. $C_3 = \{Pete, Emma, Lisa, President\} = \{5, 12, 8, 9\}$
4. $C_4 = \{Bill, Minna, Andy\} = \{4, 10, 7\}$

In order to find the peripheral nodes i.e. these that do not belong to any of the cliques, we apply the formula:

$$PN = V \setminus (C_1 \cup C_2 \cup ... \cup C_n) \tag{2}$$

where: PN — the set of peripheral nodes; V — the set of all nodes in a network; n — number of cliques within a given network

After utilizing the above formula the set NP is obtained:

$$NP = \{Mary, Rose, Mike, Peg\} = \{11, 13, 14, 15\}$$

Next part of the experiments is to identify the peripheral cliques. Let us remind that the peripheral clique (PC) is the clique that does not posses even one common node with all other cliques existing within the network. It means that each clique C_x is a PC if and only if the following criterion is met:

$$\sum_{i=1 \wedge i \neq x}^{n} card(C_i \cap C_x) = 0 \tag{3}$$

After the application of the above criterion we obtain:
$C_1 \cap C_2 = \{Ann, Lisa, Pete\} = \{1, 8, 5\}$
$C_1 \cap C_3 = \{Lisa, Pete\} = \{8, 5\}$
$C_1 \cap C_4 = \emptyset$

It does not meet the above criterion, i.e. $\sum_{i=1 \wedge i \neq 1}^{4} card(C_i \cap C_1) \neq 0$ so neither C_1 nor C_2 is not the peripheral one.
$C_2 \cap C_1 = \{Ann, Lisa, Pete\} = \{1, 8, 5\}$
$C_2 \cap C_3 = \{President, Lisa, Pete\} = \{9, 8, 5\}$
$C_2 \cap C_4 = \emptyset$

This also does not meet the above criterion: $\sum_{i=1 \wedge i \neq 2}^{4} card(C_i \cap C_2) \neq 0$ It means the product of sets is not empty, so C_2 is not the peripheral one.
$C_3 \cap C_1 = \{Lisa, Pete\} = \{8, 5\}$
$C_3 \cap C_2 = \{President, Lisa, Pete\} = \{9, 8, 5\}$
$C_3 \cap C_4 = \emptyset$
It does not meet the above criterion, i.e. $\sum_{i=1 \wedge i \neq 3}^{4} card(C_i \cap C_3) \neq 0$ $C_4 \cap C_1 = \emptyset$
$C_4 \cap C_2 = \emptyset$
$C_4 \cap C_3 = \emptyset$
$\sum_{i=1 \wedge i \neq 4}^{4} card(C_i \cap C_4) = 0$, so clique C_4 according to Formula 3 is the peripheral clique. From now on C_4 clique is a peripheral clique and is excluded from the list of cliques. The groups that remain in the set of cliques are named from now on as regular cliques.

After the identification of regular cliques as well as peripheral nodes and peripheral cliques (see Table 2) the bridges can be uncovered.

Table 2. The values of commitment function within the Thurman network

Regular Cliques	Peripheral Nodes	Peripheral Cliques
$C_1 = \{1, 2, 8, 3, 6, 5\}$ $C_2 = \{1, 8, 5, 9\}$ $C_3 = \{5, 12, 8, 9\}$	$PN = \{11, 13, 14, 15\}$	$PC = C_4 = \{4, 10, 7\}$

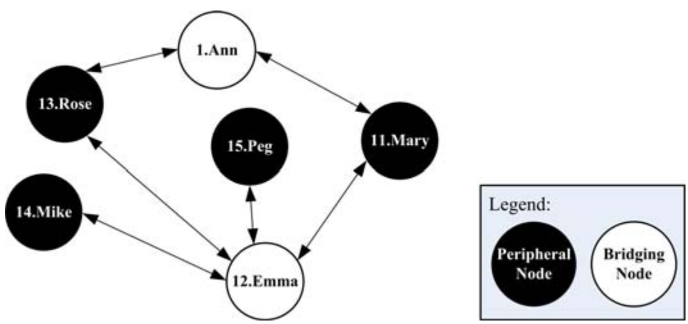

Fig. 2. Structure of peripheral nodes in Thurman network and bridges that connect *PN* with the regular cliques

In the first step the bridges that connect the peripheral nodes with the regular cliques are uncovered. In order to perform this, all of the peripheral node's connections are analyzed (Figure 2). It can be easily noticed that Ann and Emma

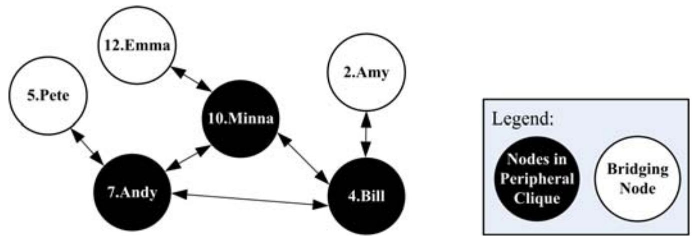

Fig. 3. Structure of peripheral clique in Thurman network and bridges that connect *PC* with the regular cliques

are bridges in this case, i.e. a set of bridges B_1 equals: $B_1 = \{Ann, Emma\} = \{1, 12\}$. Next, the analysis of peripheral clique and its relationships with the external network (Figure 3) shows that the bridging nodes in this case are $B_2 = \{Emma, Pete, Amy\} = \{12, 5, 2\}$ The final set of bridges B is the sum of the sets B_1 and B_2, so finally the set of bridges consist of:

$$B = \{Ann, Emma, Pete, Amy\} = \{1, 12, 5, 2\}$$

3.2 Properties of Bridges in Social Network

The goal of the next part of the experiments is to investigate the characteristic features of the extracted bridging nodes. For all of the nodes their social position is calculated as well as the number of mutual, incoming and outgoing edges. Before starting this part of the experiments some primary assumptions were made. The initial node positions $NP_0(x) = 1$ are established for every member x in the network. The value of ε is 0.9 and the stop condition is: no difference in node position values to the precision of 5 digits after the decimal point for all the members in two following iterations, i.e. $\tau = 0.00001$. The calculated values are presented in Table 3. Note, that bridges identified in the first part of experiments bridges posses high social positions and have the biggest number of connections — Emma, Ann, Pete and Amy. Emma has the highest social position and she is the only community member that connects both peripheral nodes (all of them) and the peripheral clique with the regular cliques. Two of the peripheral users (Peg and Mike) communicate only with Emma so she is an crucial node when concerning the cohesion of the whole network. Ann, who has the second highest social position and has 8 mutual connections, binds two peripheral users (Mary and Rose) with the regular cliques. However, these connections are not as crucial as relations between Emma and Peg or Mike, because not only Ann is connected with Mary and Rose. Another person that is connected with Mary and Rose is Emma. Pete, who obtained third highest social position, has also 8 mutual connections and is one of three people that connects the peripheral clique ({Minna, Bill, Andy}) with the regular cliques. Another users who connect the peripheral clique with the regular cliques are Emma and

Table 3. The social position values and number of edges in Thurman network

ID	Member	*SP* Ranking	*SP*	No. of mutual edges	No. of incoming edges	No. of outgoing nodes
12	Emma	1	1.90025	8	0	0
1	Ann	2	1.56732	8	0	3
5	Pete	3	1.48140	8	0	5
2	Amy	4	1.38236	6	2	0
8	Lisa	5	1.36532	7	1	0
6	Tina	6	1.17424	5	2	0
3	Katy	7	1.01320	5	1	0
10	Minna	8	0.86255	3	2	2
4	Bill	9	0.79626	3	2	0
9	President	10	0.73712	4	0	9
7	Andy	11	0.63676	3	1	0
11	Mary	12	0.60897	2	2	0
13	Rose	12	0.60897	2	2	0
14	Mike	14	0.43264	1	2	0
15	Peg	14	0.43264	1	2	0

Amy. Amy was identified also as a bridging node and she obtained the fourth position in the *SP* ranking. Additionally, she possesses 6 mutual relationships - 2 relations less than the three other bridging nodes. On the other hand President, who has many connections but low social position (10th in *SP* ranking), is not a bridging node. It means that social position is better measure to identify the bridges than the number of the edges.

4 Conclusions

The method of detecting bridging nodes in a directed and weighted social network as well as their properties were investigated. The conducted research reveals that the bridges obtain the highest social position. Two types of bridges can be distinguished: (i) these that connect peripheral nodes with regular cliques and (ii) these that connect peripheral cliques with the regular cliques. Note that the highest social position is obtained by Emma (12) who binds both all of the peripheral nodes and peripheral clique with the regular cliques. Ann (1), Pete (5) and Amy (2) also are the bridging nodes and all of them obtain high social positions: Ann — second, Pete — third and Amy — fourth. This reveals that social position is a good measure that can be used to find the bridging nodes within the network. In other words, there exists the correlation between the social position of the network and the fact if the node is a bridging one. The future work will focus on conducting the research within the complex social networks with big number of nodes and edges. We intend to develop a fast method of identification of the bridge nodes then to use it to track changes in community dynamics (group evolution). The changes in the properties of bridging nodes may be used to track processes of merging, splitting, growth and contraction of cliques in complex social networks.

Acknowledgment. The work was supported by The Polish Ministry of Science and Higher Education, grant no. N N516 264935.

References

1. Burt, R.S.: Structural holes. Harvard University Press, Cambridge (1992)
2. Carrington, P., Scott, J., Wasserman, S.: Models and methods in Social Network Analysis. Cambrige University Press, Cambrige (2005)
3. Granovetter, M.S.: The strength of weak ties: A Network Theory Revisited. Sociological Theory 1, 201–233 (1983)
4. Hanneman, R., Riddle, M.: Introduction to social network methods. Online textbook, `http://faculty.ucr.edu/~hanneman/nettext/` (April 1, 2006)
5. Harary, F., Norman, R.Z., Cartwright, D.: Structural Models: An Introduction to the Theory of Directed Graphs. John Wiley and Sons, New York (1965)
6. Hatala, J.P.: Social Network Analysis in Human Resources Development: A New Methodology. Human Resource Development Review 5(1), 45–71 (2006)
7. Kazienko, P., Musiał, K.: Assessment of Personal Importance Based on Social Networks. In: Gelbukh, A., Kuri Morales, Á.F. (eds.) MICAI 2007. LNCS (LNAI), vol. 4827, pp. 529–539. Springer, Heidelberg (2007)
8. Kazienko, P., Musiał, K., Zgrzywa, A.: Evaluation of Node Position Based on Email Communication. Control and Cybernetics (in press, 2009)
9. Luce, R.D., Perry, A.D.: A method of matrix analysis of group structure. Psychometrica 14, 95–116 (1949)
10. Scott, J.: Social network analysis: A handbook, 2nd edn. Sage, London (2000)
11. Thurman, B.: In the office: Networks and coalitions. Social Networks 2, 47–63 (1979)
12. Wasserman, S., Faust, K.: Social network analysis: Methods and applications. Cambridge University Press, New York (1994)