Chapter 6
Online Interactions

Lilian Weng, Filippo Menczer, and Alessandro Flammini

Abstract The ubiquitous use of the Internet has led to the emergence of countless social media and social networking platforms, which generate large-scale digital data records of human behaviors online. Here we review the literature on online interactions, focusing on two main themes: social link formation and online communication. The former is often studied in the context of network evolution models and link prediction or recommendation tasks; the latter combines classic social science theories on collective human behaviors with analysis of big data enabled by advanced computation techniques. But the structure of the network, and the flow of information through the network influence each other. We present a case study to illustrate the connections between social link formation and online communication. Analysis of longitudinal micro-blogging data reveals that people tend to follow others after seeing many messages by them. We believe that research on online interactions will benefit from a deeper understanding of the mutual interactions between the dynamics *on* the network (communication) and the dynamics *of* the network (evolution).

6.1 Introduction

User activity within online socio-technical systems is exploding. Social and micro-blogging networks such as Facebook, Twitter, and Google Plus host the information sharing activity of billions of users every day. Using these network platforms, people communicate ideas, opinions, videos, and photos among their circles of friends and followers across the world. These interactions generate an unprecedented amount of data that can be used as a social observatory, providing a unique opportunity to study the mechanisms behind human interactions with a quantitative approach [1–4].

Research on human online interactions revolves around two main themes: social link formation and online communication. People can build virtual connections with others to subscribe to their messages (i.e., *following* on Twitter and Google

L. Weng (✉) • F. Menczer • A. Flammini
Center for Complex Networks and Systems Research, School of Informatics and Computing,
Indiana University Bloomington, Bloomington, IN, USA
e-mail: weng@umail.iu.edu

Plus) or to claim a mutual friendships between them (i.e., *friending* on Facebook). The established social links enable people to easily communicate with friends, sharing and spreading information on top of the social network.

Most work on social link formation frames the problem as network evolution modeling, in which each new link creation is driven by predefined mechanisms that resemble the observations from real-world data. Meanwhile, many algorithms have been presented to predict or recommend missing links in a given network.

Communication dynamics is a long-lasting research topic in the social sciences. Many theories of how people interact, exchange ideas, and influence each other have been proposed decades ago. With the availability of big data and advanced computational power, researchers can apply, verify, and enrich classic hypotheses on human behaviors, leveraging the capacity to collect and analyze data on a large scale to reveal patterns of human interactions [3]. New interdisciplinary research fields, namely *computational social science* and *human dynamics*, have emerged in such a scenario [3, 5].

6.1.1 Social Link Formation

Understanding the formation of online social links is a key ingredient for modeling the evolution of online social networks, as the rules for creating new links determine the network structure in time. Various models were introduced to capture the growth and evolution of network topology, as well as different characteristics of complex networks. Most such models focused on defining basic mechanisms that drive link creation [6–10].

6.1.1.1 Classic Network Evolution Models

The random network model is the oldest attempt at characterizing a non-regular network. Although, strictly speaking, it is not an evolutionary model, it can be regarded as such when links are added sequentially to the network. It is characterized by the fact that each link exists independently with the same probability [11]. The random network inspired many subsequent studies in network science, but it was not thought to reproduce several crucial properties of social networks [12], such as the small-world effect [13–15], high clustering coefficient [14, 16, 17], temporal dynamics [18, 19], information propagation [20], and heterogeneous distributions in connectivity patterns [21–25]. Many such characteristics were indeed yet undiscovered at the time when the random network model was proposed.

The small-world effect, also known as "six degrees of separation," originated from the Milgram experiment [13], in which the average length of communication chains between two random individuals was found to be around six—smaller than

what expected in a regular network, such as a lattice. The Watts-Strogatz model was designed to reproduce the small-world phenomenon by rewiring each link in a regular network with a small probability [14].

A scale-free network has a power-law degree distribution, commonly seen in many real-world social networks such as the film actor network, the scientific collaboration network, and the citation network [6, 8, 21]. A highly skewed degree distribution in a social network indicates that, although the majority of nodes is poorly connected, there is a consistent group of them (compared to what happens in a random graph) that is extremely well connected, and whose collective connections account for a relevant portion of the entire set of links in the network. In a large group of people, only a few are extremely popular and most others do not have many contacts. Many models have been proposed to reproduce the heterogeneous distribution in connectivity [12, 22–25]. The Barabási-Albert model generates a scale-free network by continuously adding new nodes into the system ("growth") and connecting them with other nodes with preference to high-degree ones ("preferential attachment") [21]. Motivated by the structure of the Web graph, the copying model adds a new node into the network at a time and links it to a random existing node or its neighbors [24, 25]. Another model proposed by Newman, Watts, and Strogatz aimed to build up a random graph with arbitrary degree settings [12]. The *ranking model* grows the network according to a rank of the nodes by any given prestige measure, reproducing arbitrary power-law degree distributions [23].

6.1.1.2   Models with Social Components

The preferential attachment mechanisms in the Barabási-Albert and ranking models have a clear rationale in the social context: people prefer to form edges with well-connected individuals, such as celebrities. However, this prescription alone is not sufficient to reproduce several other important features of real networks. Other models have been put forth to fill the gap, including ingredients such as homophily [26–29], triadic closure [15–17, 30, 31], hierarchical structure [32], and information diffusion [20, 33].

Homophily can be regarded as people's propensity for linking with similar others [26, 28, 29]. The triadic closure mechanism is based on the intuition that two individuals with mutual friends have a higher probability to establish a new contact [30, 31]. This tendency was observed in both undirected and directed online social networks and incorporated into several network growth models [15–17]. In particular, Leskovec et al. tested triadic closure against many other mechanisms in four different large-scale social networks. By using maximum likelihood estimation (MLE) [34], they identified triadic closure as the best rule, among those considered, to explain link creation and to reproduce the clustering coefficient and the degree distribution of the real networks under study [16].

6.1.1.3   Link Prediction

Missing link prediction algorithms aim at inferring new social connections that may happen in the near future given a current snapshot of the network structure. The prediction has practical applications in the online systems; one of the most popular use cases is to provide recommendations for new contacts (i.e., "People You May Know" on Facebook and LinkedIn). Common approaches consider link prediction as a classification task or ranking problem, using node similarity [35, 36], the hierarchical structure of the network [32], random walks [16], supervised random walks [37], graphical models [38], and user profile features [39].

One class of link prediction algorithms is designed on the premise that only the network topology is known. Liben-Nowell and Kleinberg [36] examined and compared a rich set of metrics for quantifying the similarity between a pair of nodes in the social network, where high similarity implies high likelihood of being connected on the basis of homophily. Tested metrics were built upon different network topological features associated with each node, including overlap between neighbor sets, preferential attachment, shortest path distance, and PageRank hitting time. The analysis identified Adamic-Adar similarity [35] as the metric providing the best performance. Clauset, Moore, and Newman [32] explored the observation that real-world networks often exhibit community structure and hierarchical organization. They proposed a link prediction method that uses knowledge about the network's hierarchical structure.

Other link prediction algorithms depend on additional attributes of existing individuals or connections, as well as the topology of the network. Supervised random walks can incorporate the knowledge of nodes and links so that a random walker is guided to follow preferred paths with higher probability [37]. Attributes may include, for example, the number of co-authored papers in a collaboration network and the frequency of interactions between a pair of friends on Facebook. In online systems like Flickr and Last.fm, users can annotate content revealing their topical interests. Schifanella et al. found that users with similar interests in these networks are more likely to be friends and proposed to use the similarity between user annotation metadata as a predictor of missing social links [39].

6.1.2   Communication Dynamics

Early models concerning communication dynamics were inspired by studies of epidemics, assuming that a piece of information could pass from one individual to another through social contacts [40–42]. Recently, starting from observations and theories in social sciences, a wealth of computational models have been proposed to describe human communication.

6.1.2.1 Threshold Model

One class of models is based on the idea of a threshold: people tend to follow the same trends as most of their friends do [43, 44]. A threshold can be defined as the number or fraction of others who must make a decision before a given actor does the same. Many empirical studies have demonstrated the existence of such a threshold in social and behavioral contagion online [45–48]. Threshold models have been widely applied to understand the diffusion of rumors, norms, strikes, voting, educational attainment, migration, and other human behaviors [43, 49, 50], and extended to study the role of competition for finite attention [51].

6.1.2.2 Homophily

The principle of homophily states that similar people are more likely to have contact than dissimilar ones [27, 52, 53]. The existence of homophily in social groups has been supported by various empirical observations and experiments in online settings [26, 29, 39, 54, 55]. Crandall et al. proposed a homophily-based model to predict a user's future activity and interactions with others according to user similarities [56].

A feedback loop has been claimed to result in increasing similarity among users: people grow to resemble their friends because of social (peer) influence, while being more likely to form links with similar people (homophily) [29, 56]. Such a feedback loop could lead to the so-called *echo-chamber* effect, by which people are exposed to limited diversity of opinions in online social networks [57, 58]. Though it is hard to fully distinguish between peer influence and homophily [59], the latter effect contributes to promoting behavioral contagion [54].

6.1.2.3 Weak Tie Hypothesis

Friendships vary in their intensity and intimacy. The concept of tie strength has been introduced to capture this variation: strong ties are our closest confidants and supporters, while weak ties, to whom we feel less close, comprise the majority of our personal networks. Granovetter defined the strength of social ties proportionally to the size of shared social circles and proposed the weak-tie hypothesis [31, 60], according to which weak ties do not carry as much communication as strong ties, but act as bridges between communities and thus as important channels for novel information.

Following up on Granovetter's work, many empirical studies have tested the weak-tie hypothesis [61–69]. Brown and Reingen found an important bridging function of weak ties in word-of-month referral behavior, allowing information to travel from one distinct subgroup of referral actors to another [62]. Gilbert and Karahalios tested several dimensions of tie strength on social media, revealing that both intensity of communication and intimate language are strong indicators of

relationship closeness [69]. Onnela et al. analyzed a mobile call network and showed that individuals in clusters tend to communicate more, while ties between clusters have less traffic [68]. Bakshy et al. compared individual adoption rates on Facebook when an external URL shared by friends is or is not included in the newsfeed and found that although stronger ties are individually more influential in persuading others to adopt and spread information, more abundant weak ties are responsible for the propagation of novel information [61].

In summary, strong ties are believed to provide greater emotional support [69, 70] and to be more influential [61, 62, 71], while weak ties provide novel information and connect us to opportunities outside our immediate circles [31, 68, 72].

6.1.2.4 Limited Attention

People have limited attention during communication. This constraint may be related to a cognitive limit on the number of stable social relationships that one can sustain, as postulated by Dunbar [73] and later supported by analyses of Twitter data [74, 75]. Huberman, Romero, and Wu defined friends of a Twitter user as those who have been mentioned at least twice. They found that most users have a very small number of friends compared to a large number of followers, and the friend network is more influential than the follower network in driving Twitter usage [74]. Wu and Huberman analyzed the dynamics of collective attention on Digg.com and modeled the delay of collective attention with a single novelty factor. Their measurements indicated that novelty within groups decays with a stretched-exponential law, suggesting the existence of a natural time scale over which attention fades [76].

6.1.2.5 Communication Dynamics on Evolving Networks

The large majority of studies on communication dynamics consider a static underlying social network, under the assumption that the network evolves on a slower time scale than that characteristic of the information spread. Recent research has addressed the modeling of cases in which the time scales of communication dynamics and network evolution are comparable. These approaches consider the two processes as either independent [19, 77] or coupled [33, 78, 79]. In particular, the studies focused on the former case considered mainly epidemic processes in which links are deleted or rewired according to the disease status of each node [78, 79].

6.2 Case Study: Traffic-Based Social Link Formation

We probe into the effects of information diffusion in shaping the evolution of the social network structure. As a case study, we present a longitudinal analysis of micro-blogging data to better understand the strategies employed by users when

expanding their social circles. While the network structure affects the spread of information, the network is, in turn, shaped by this communication activity. This leads us to hypothesize a mechanism whereby people tend to follow others after seeing many messages from them. Interestingly, the coupling of social link formation and information sharing allows to depict a more accurate and comprehensive view of the network evolution [33].

We analyzed a dataset collected from *Yahoo! Meme*,[1] including the entire history of the system from April 2009 until March 2010. A user j following a user i is represented in the follower network by a directed edge $\ell = (i, j)$, indicating j can receive messages posted by i. We adopt this notation, in which the link creator is the target, to emphasize the direction of information flow. In our notation, the in-degree of a node j is the number of people followed by j. Users can repost received messages, which become visible to their followers. When user j reposts content from i, we infer a flow of information from i to j. Each link is weighted by the numbers of messages from i that are reposted or seen by j. At the end of the observation period, the Yahoo! Meme follower network consisted of 128,199 users with at least one edge, connected by a total of 3,485,361 directed edges.

Social micro-blogging networks, such as Twitter, Google Plus, Sina Weibo, and Yahoo! Meme, are designed for information sharing. As illustrated in Fig. 6.1, the dynamics *on* the network directly affects the dynamics *of* the networks, and vice versa. In this case study, we investigate the individual strategies that lead to the creation of new social links. We characterize link creation processes with a set of parameters associated with different link creation strategies, estimated by a Maximum-Likelihood approach [34]. This analysis will show that triadic closure does have a strong effect on link formation, but shortcuts based on traffic are another indispensable factor in interpreting network evolution.

6.2.1 Link Creation Mechanisms

When users post or repost messages, all their followers can see these posts and might decide to repost them, generating spreading paths that, when taken together, form cascade networks. When receiving a reposted message, a Meme user in such a path can see both the *grandparent* (G, the user two steps ahead in the path) and the *origin* (O, original source). A user may decide to follow a grandparent or origin, receiving their future messages directly. These new links create *shortcuts* connecting users at any distance in the network. A triadic closure occurs when a user follows a *triadic node* (Δ, the user two steps away in the follower network). The definitions of different types of link creation mechanisms are illustrated in Fig. 6.2.

---

[1]Yahoo! Meme was a social micro-blogging system similar to Twitter, active between 2009 and 2012.
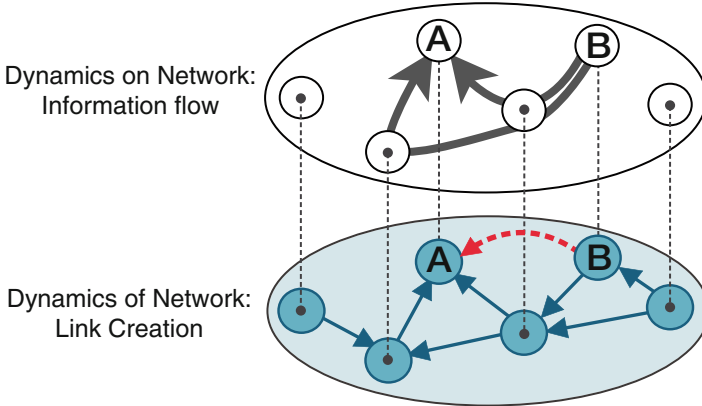
Fig. 6.1 The dynamics *of* and *on* the network are strongly coupled. The *bottom layer* illustrates the social network structure, where the *blue arrows* represent "follow" relationships with the direction of information flow. The *dashed red arrow* marks a newly created link. The *upper layer* depicts the flow of information between people in the same group, leading to the creation of the new link. The social network structure constrains communication patterns, but information propagated through the network also affect how agents behave and ultimately how the network changes and grows
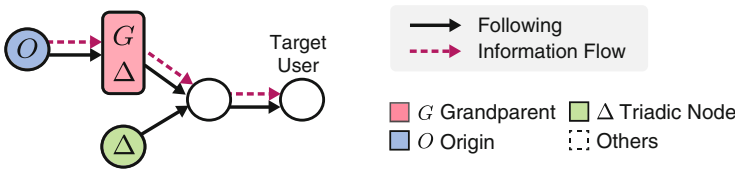

Fig. 6.2 Illustration of link creation mechanisms. A grandparent node is a special case of triadic node, from which or through which information has reached the target user. Therefore traffic-based shortcuts to grandparent nodes are a subset of triadic closures

6.2.1.1 Statistical Analyses of Shortcuts

To quantify the statistical tendency of users to create shortcuts, let us consider every single link creation in the data as an independent event. We test the null hypotheses that links to grandparents, origins, and triadic nodes are generated by choosing targets at random among the users not already followed by the creator.

We label each link ℓ by its creation order, $1 \le \ell \le L$, where L is the total number of links. For each link, we can compute the likelihood of following a grandparent by chance:

$$p_G(\ell) = \frac{N_G(\ell)}{N(\ell) - k(\ell) - 1}, \tag{6.1}$$

where $N_G(\ell)$ is the number of distinct grandparents seen by the creator of ℓ at the moment when ℓ is about to be created; $N(\ell)$ is the number of available users in the system when ℓ is to be created; $k(\ell)$ is the in-degree of ℓ's creator at the same moment; and the denominator is the number of potential candidates to be followed.

The indicator function for each link ℓ denotes whether the link connects with a grandparent or not in the real data:

$$\mathbf{1}_G(\ell) = \begin{cases} 1 \text{ if } \ell \text{ links to a grandparent in the data} \\ 0 \text{ otherwise.} \end{cases} \quad (6.2)$$

The expected number of links to grandparents according to the null hypothesis can be then computed as:

$$E_G = \sum_{\ell=1}^{L} p_G(\ell) \quad (6.3)$$

and its variance is given by:

$$\sigma_G^2 = \sum_{\ell=1}^{L} p_G(\ell)(1 - p_G(\ell)) \quad (6.4)$$

while the corresponding empirical number is:

$$S_G = \sum_{\ell=1}^{L} \mathbf{1}_G(\ell). \quad (6.5)$$

According to the Lyapunov central limit theorem,[2] the variable $z_G = (S_G - E_G)/\sigma_G$ is distributed according to a standard normal $\mathcal{N}(0, 1)$. For linking to origins (O) or triadic nodes (Δ), we can define z_O and z_Δ similarly. In all three cases, using a z-test, we can reject the null hypotheses with high confidence ($p < 10^{-10}$). We conclude that links established by following grandparents, origins or triadic nodes happen much more frequently than by random connection. These link creation mechanisms have important roles in the evolution of the social network.

---

[2]Lyapunov's condition, $\frac{1}{\sigma_n^4} \sum_{\ell=1}^{n} E[(X(\ell) - p(\ell))^4] \xrightarrow{n \to \infty} 0$ where $X(\ell)$ is a random Bernoulli variable with success probability $p(\ell)$ [80], is consistent with numerical tests. Details are omitted for brevity.

Fig. 6.3 Individual preferences for following grandparents (*red circles*), origins (*blue squares*), and triadic nodes (*green triangles*) change with the in-degree of the link creator
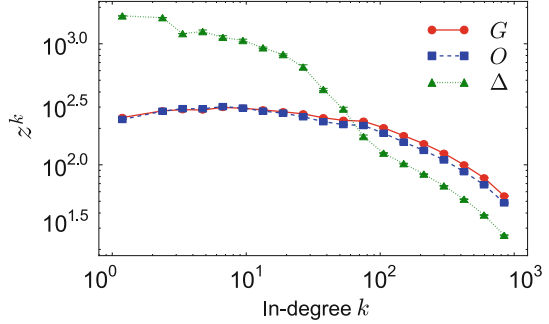
6.2.1.2 User Preference

The variables z_G, z_O, and z_Δ, as defined above, measure how much more likely links of a given type are formed than by chance—in other words, how strong individual preferences are for following grandparents, origins or triadic nodes. To study the dependence of the link formation tendencies on the different stages of an individual's lifetime, let us compute z_G^k, z_O^k, and z_Δ^k for links created by users with in-degree k, that is, those who are following k users at the time when the link is created. Figure 6.3 shows that the principle of triadic closure dominates user behavior when one follows a small number of users ($k < 75$). In the early stages, one does not receive much traffic, so it is natural to follow people based on local social circles, consistently with triadic closure. However, users who have been active for a long time and have followed many people ($k > 75$) have more channels through which they monitor traffic. This creates an opportunity to follow others from whom they have seen messages in the past.

6.2.1.3 Link Efficiency

In information diffusion networks like Twitter and Yahoo! Meme, social links may have a key efficiency function of shortening the distance between information creators and consumers. An efficient link should be able to convey more information to the follower compared to less efficient links. Hence we define the *efficiency* of link ℓ as the average number of posts seen or reposted through ℓ during one time unit after its creation:

$$\eta_{\text{seen}} = \frac{w_{\text{seen}}(\ell)}{T - t(\ell)}, \quad \eta_{\text{repost}} = \frac{w_{\text{repost}}(\ell)}{T - t(\ell)}, \quad (6.6)$$

where $w(\ell)$ is the number of messages seen or reposted through ℓ; $t(\ell)$ is the time when ℓ was created; and T is the time of the last action recorded in our dataset. Both seen and reposted messages are considered, as they represent different types of traffic; the former are what is visible to a user, and the latter are what a user is
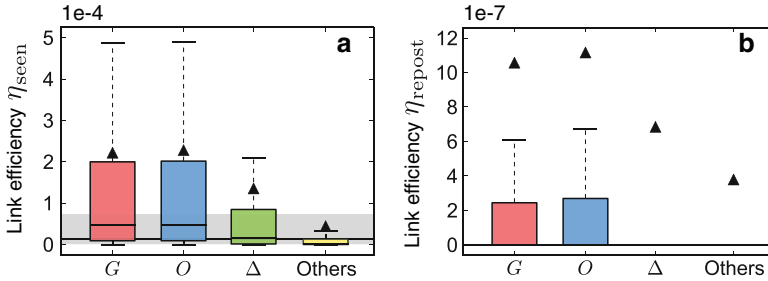
Fig. 6.4 Efficiency of links created according to different mechanisms, or average number of messages (**a**) seen or (**b**) reposted per time unit. Each *box* shows data within lower and upper quartile. Whiskers represent the 99th percentile. The *triangle* and *line* in a box represent the mean and median, respectively. The *black line* and *grey area* across the entire figure mark the median and interquartile range of the measure across all links, respectively

willing to share. We compute the link efficiency of every grandparent, origin, and triadic closure link. As shown in Fig. 6.4, both grandparent and origin links exhibit higher efficiency than triadic closure links, irrespective of the type of traffic. By shortening the paths of information flows, more posts from the content generators reach the consumers.

6.2.2 Rules of Network Evolution

To infer the different link creation strategies from the observed data, we characterize users with a set of probabilities associated with different actions, and approximate these parameters by MLE [34]. For each link ℓ, we know the actual creator and the target; we can thus compute the likelihood $f(\ell|\Gamma, \Theta)$ of the target being followed by the creator according to a particular strategy Γ, given the network configuration Θ at the time when ℓ is created. The likelihoods associated with different strategies can be mixed according to the parameters to obtain a model of link creation behavior. Finally, assuming that link creation events are independent, we can derive the likelihood of obtaining the empirical network from the model by the product of likelihoods associated with every link. The higher the value of the likelihood function, the more *accurate* the model.

6.2.2.1 Simple Strategy

We consider five link creation mechanisms and their combinations:

- *Random* (Rand): follow a randomly selected user who is not yet followed
- *Triadic closure* (Δ): follow a randomly selected triadic node
- *Grandparent* (G): follow a randomly selected grandparent

- *Origin* (*O*): follow a randomly selected origin
- *Traffic shortcut* (*G* ∪ *O*): follow a randomly selected grandparent or origin

Other mechanisms for link creation could be similarly incorporated, such as social balance [81] and preferential attachment [21]. However, preferential attachment is built on the assumption that everyone knows the global connectivity of everyone else, which is not very realistic. The strategies considered here essentially reproduce and extend the copy model [25], approximating preferential attachment with only local knowledge.

To model link creation with a single strategy, we can use a parameter p for the probability of using that strategy, while a random user is followed with probability $1 - p$. The calculation of maximum likelihood, taking the single strategy of grandparents as an example, is as follows:

$$\mathscr{L}_G(p) = \prod_{\ell=1}^{L} (p f(\ell|G, \Theta) + (1-p) f(\ell|\text{Rand}, \Theta))$$

$$= \prod_{\ell=1}^{L} \left(p \frac{\mathbf{1}_G(\ell)}{N_G(\ell)} + (1-p) \frac{1}{N(\ell) - k(\ell) - 1} \right)$$

$$= \prod_{\mathbf{1}_G(\ell)=1} \left(\frac{p}{N_G(\ell)} + \frac{1-p}{N(\ell) - k(\ell) - 1} \right) \prod_{\mathbf{1}_G(\ell)=0} \frac{1-p}{N(\ell) - k(\ell) - 1}. \quad (6.7)$$

Note that since a follow action can be ascribed to multiple strategies, it can contribute to multiple terms in the log-likelihood expression. For instance, a link could be counted in both $f(\ell|G, \Theta)$ and $f(\ell|\text{Rand}, \Theta)$. For numerically stable computation, we maximize the log-likelihood:

$$\log \mathscr{L}_G(p) = \sum_{\mathbf{1}_G(\ell)=1} \ln \left(\frac{p}{N_G(\ell)} + \frac{1-p}{N(\ell) - k(\ell) - 1} \right)$$

$$+ \sum_{\mathbf{1}_G(\ell)=0} \ln \frac{1-p}{N(\ell) - k(\ell) - 1}. \quad (6.8)$$

Similar expressions of log-likelihood can be obtained for other strategies (Δ, O, and $G \cup O$).

It is not trivial to obtain the best p analytically, so we explore the values of $p \in (0, 1)$ numerically (see Fig. 6.5). Triadic closure dominates as a single strategy, with $p_\Delta = 0.82$. Traffic-based strategies alone account for about 20 % of the links.
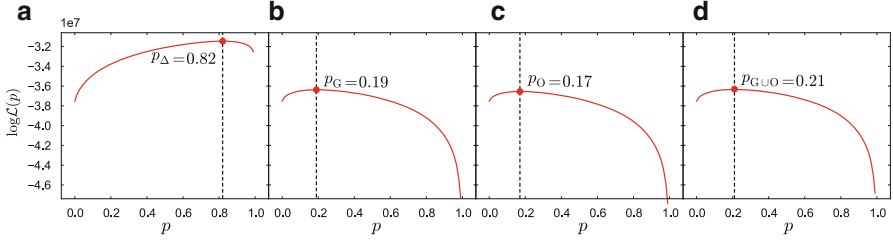
Fig. 6.5 The plot of the log-likelihood $\log \mathscr{L}(p)$ as a function of link creation strategy probabilities for models with a single strategy. The *red circles* mark the maximized $\log \mathscr{L}(p)$. (**a**) Triadic closure, (**b**) Grandfather, (**c**) Origin, (**d**) Grandfather + Origin

6.2.2.2   Combined Strategies

For a more realistic model of the empirical data, let us consider combined strategies with both triadic closure and traffic-based shortcuts. For each link ℓ, the follower with probability p_1 creates a shortcut by linking to a grandparent or an origin ($G \cup O$); with probability p_2 follows a triadic node (Δ); and with probability $1 - p_1 - p_2$ connects to a random node. Taking the combined strategy with grandparent as an example, we compute the log-likelihood as:

$$\log \mathscr{L}_{G+\Delta}(p_1, p_2) = \log \prod_{\ell=1}^{L} \Big[p_1 f(\ell | G \cup O, \Theta) + p_2 f(\ell | \Delta, \Theta)$$

$$+ (1 - p_1 - p_2) f(\ell | \text{Rand}, \Theta) \Big]. \tag{6.9}$$

The detailed derivations of the likelihood functions and the cases of the other combined strategies are omitted for brevity.

Once again, we numerically explore the values of p_1 and p_2 in the unit square to maximize the log-likelihood. The likelihood landscape for the combined strategy considering both grandparents and origins as well as triadic closure is shown in Fig. 6.6. The parameter settings and the maximum likelihood values for all tested models are listed in Table 6.1. We can compare the quality of these models by comparing their maximized $\log \mathscr{L}$'s. The combined models with both traffic shortcuts and triadic closure yield the best accuracy. In these models, triadic closure accounts for 71 % of the links, grandparents and origins for 12 %, and the rest of the links are created at random.

Fig. 6.6 The contour plot of
log-likelihood $\log \mathscr{L}(p_1, p_2)$
for the combined strategy of
creating traffic shortcuts
$(G \cup O)$ with probability p_1
and triadic closure links (Δ)
with probability p_2. The *black
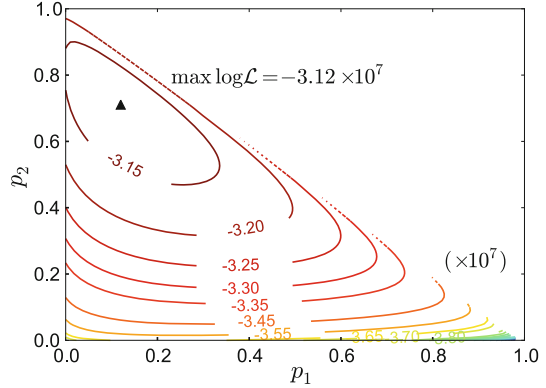triangle* marks the optimum

Table 6.1 The best parameters in different models and corresponding values of maximized log-likelihood function

Strategy	Model	Parameters		max $\log \mathscr{L}$
Single	Δ	$p = 0.82$		-3.15×10^7
	G	$p = 0.19$		-3.64×10^7
	O	$p = 0.17$		-3.65×10^7
	$G \cup O$	$p = 0.21$		-3.63×10^7
Combined	$G + \Delta$	$p_1 = 0.12$	$p_2 = 0.71$	-3.12×10^7
	$O + \Delta$	$p_1 = 0.10$	$p_2 = 0.73$	-3.13×10^7
	$G \cup O + \Delta$	$p_1 = 0.12$	$p_2 = 0.71$	-3.12×10^7

6.3 Discussion

Social link formation and information sharing are two major tracks of research on
online interactions. The mechanisms of new link creation determine the topology
of linkages among individuals, and the underlying network structure is critical for
the dynamics of the diffusion process [6–8, 51]. At the same time, as many social
links are driven by the need for more efficient information sharing in social media
sites, social link formation is greatly affected by communication activity. Both the
evolving structure of the social network and information diffusion have been studied
for decades, but the coupling between these dynamical processes has not been
well explored. In the present case study, we demonstrate a feedback loop between
these two dynamics. While triadic closure is the dominant mechanism for social
network evolution, it is mainly relevant in the early stages of a user's lifetime. As
time progresses, the traffic generated by communication dynamics on the network
becomes an indispensable component for user linking behavior. As users become
more active and influential, their links create shortcuts that make the spread of
information more efficient in the network.

Studies of online interactions—how social networks evolve and how information
spreads—help us gain a better understanding of social influence, user behavior, and

network efficiency in the context of online systems. The coupling between dynamics *of* and *on* the network provides us with powerful insights into human interactions in the digital world.

References

1. Cho, A. (2009). Ourselves and our interactions: The ultimate physics problem? *Science, 325*, 406.
2. Kumar, R., Novak, J., & Tomkins, A. (2006). Structure and evolution of online social networks. In *Proceedings of SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM.
3. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A. -L., Brewer, D., et al. (2009). Computational social science. *Science, 323*(5915), 721–723.
4. Vespignani, A. (2009). Predicting the behavior of techno-social systems. *Science, 325*(5939), 425–428.
5. Barabási, A. -L., & Albert, R. (2005). The origin of bursts and heavy tails in human dynamics. *Nature, 435*(7039), 207–211.
6. Albert, R., & Barabási, A. -L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics, 74*(1), 47–97.
7. Barrat, A., Barthélemy, M., & Vespignani, A. (2008). *Dynamical processes on complex networks*. Cambridge: Cambridge University Press.
8. Newman, M. E. J. (2003). The structure and function of complex networks. *SIAM Review, 45*(2), 167–256.
9. Newman, M. E. J. (2010). *Networks: An introduction*. Oxford: Oxford University Press.
10. Wasserman, S., & Faust, K. (1994). *Social network analysis*. Cambridge: Cambridge University Press.
11. Erdös, P., & Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5*, 17–61.
12. Newman, M. E. J., Watts, D. J., & Strogatz, S. H. (2002). Random graph models of social networks. *Proceedings of the National Academy of Sciences (PNAS), 99*(Suppl 1), 2566–2572.
13. Milgram, S. (1967). The small world problem. *Psychology Today, 2*(1), 60–67.
14. Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440–442.
15. Krackhardt, D., & Handcock, M. S. (2007). Heider vs. simmel: Emergent features in dynamic structure. In E. M. Airoldi, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, A. X. Zheng (Eds.), *Statistical network analysis: models, issues, and new directions* (pp. 14–27). Berlin: Springer.
16. Leskovec, J., Backstrom, L., Kumar, R., & Tomkins, A. (2008). Microscopic evolution of social networks. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 462–470).
17. Romero, D. M., & Kleinberg, J. (2010). The directed closure process in hybrid social-information networks, with an analysis of link formation on Twitter. In *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*.

18. Perra, N., Gonçalves, B., Pastor-Satorras, R., & Vespignani, A. (2012). Time scales and dynamical processes in activity driven networks. *Nature Scientific Reports, 2*, 469.
19. Rocha, L. E. C., Liljeros, F., & Holme, P. (2011). Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Computational Biology, 7*(3), e1001109.
20. Barbieri, N., Bonchi, F., & Manco, G. (2013). Cascade-based community detection. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)* (pp. 33–42).
21. Barabási, A. -L., & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286*(5439), 509–512.
22. Dorogovtsev, S., Mendes, J., & Samukhin, A. (2000). Structure of growing networks with preferential linking. *Physical Review Letters, 85*(21), 4633–4636.
23. Fortunato, S., Flammini, A., & Menczer, F. (2006). Scale-free network growth by ranking. *Physical Review Letters, 96*(21), 218701.
24. Kleinberg, J., Kumar, R., Raghavan, P., Rajagopalan, S., & Tomkins, A. (1999). The web as a graph: measurements, models and methods. *Lecture Notes in Computer Science (LNCS), 1627*, 1–18.
25. Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., & Upfal, E. (2000). Stochastic models for the web graph. In *Proceedings of IEEE Annual Symposium on Foundations of Computer Science* (pp. 57–65).
26. Gallos, L., Rybski, D., Liljeros, F., Havlin, S., & Makse, H. (2012). How people interact in evolving online affiliation networks. *Physical Review X, 2*(3), 031014.
27. McPherson, M., Lovin, L., & Cook, J. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444.
28. Papadopoulos, F., Kitsak, M., Ángeles Serrano, M., Boguña, M., & Krioukov, D. (2012). Popularity versus similarity in growing networks. *Nature, 489*(7417), 537–540.
29. Weng, L., & Lento, T. (2014). Topic-based clusters in egocentric networks on Facebook. In *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*.
30. Simmel, G., & Wolff, K. H. (1950). *The Sociology of Georg Simmel*. New York: The Free Press.
31. Granovetter, M. (1973). The strength of weak ties. *American Journal of Sociology, 78*(6), 1.
32. Clauset, A., Moore, C., & Newman, M. (2008). Hierarchical structure and the prediction of missing links in networks. *Nature, 453*(1), 98–101.
33. Weng, L., Ratkiewicz, J., Perra, N., Gonçalves, B., Castillo, C., Bonchi, F., et al. (2013). The role of information diffusion in the evolution of social networks. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 356–364).
34. Cowan, G. (1998). *Statistical data analysis*. Oxford: Oxford Science Publications.
35. Adamic, L. A., & Adar, E. (2003). Friends and neighbors on the web. *Social Networks, 25*(3), 211–230.
36. Liben-Nowell, D., & Kleinberg, J. (2007). The link-prediction problem for social networks. *Journal of American Society for Information Science and Technology (JASIST), 58*(7), 1019–1031.
37. Backstrom, L., & Leskovec, J. (2011). Supervised random walks: Predicting and recommending links in social networks. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)* (pp. 635–644). ACM.
38. Lou, T., Tang, J., Hopcroft, J., Fang, Z., & Ding, X. (2010). Learning to predict reciprocity and triadic closure in social networks. *ACM Transactions on Embedded Computing Systems, 9*(4), 5.
39. Schifanella, R., Barrat, A., Cattuto, C., Markines, B., & Menczer, F. (2010). Folks in folksonomies: social link prediction from shared metadata. In *Proceedings of ACM International Conference on Web Search and Data Mining (WSDM)* (pp. 271–280).
40. Anderson, R. M., May, R. M., & Anderson, B. (1992). *Infectious diseases of humans: Dynamics and control* (Vol. 28). Oxford: Oxford University Press.

41. Daley, D. J., & Kendall, D. G. (1964). Epidemics and rumours. *Nature, 204*(4963), 1118–1119.
42. Goffman, W., & Newill, V. A. (1964). Generalization of epidemic theory: An application to the transmission of ideas. *Nature, 204*(4955), 225–228.
43. Granovetter, M. S. (1978). Threshold models of collective behavior. *American Journal of Sociology, 83*(6), 1420–1433.
44. Morris, S. (2000). Contagion. *Review of Economic Studies, 67*(1), 57–78.
45. Backstrom, L., Huttenlocher, D., Kleinberg, J., & Lan, X. (2006). Group formation in large social networks: Membership, growth, and evolution. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 44–54).
46. Bakshy, E., Karrer, B., & Adamic, L. (2009). Social influence and the diffusion of user-created content. In *Proceedings of ACM Conference on Electronic Commerce* (pp. 325–334).
47. Cosley, D., Huttenlocher, D., Kleinberg, J., Lan, X., & Suri, S. (2010). Sequential influence models in social networks. In *Proceedings of AAAI International Conference on Weblogs and Social Media (ICWSM)*.
48. Romero, D. M., Meeder, B., & Kleinberg, J. (2011). Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on twitter. In *Proceedings of International Conference on World Wide Web (WWW)*.
49. Centola, D. (2010). The spread of behavior in an online social network experiment. *Science, 329*(5996), 1194–1197.
50. Weng, L., Menczer, F., & Ahn, Y. -Y. (2013). Virality prediction and community structure in social networks. *Nature Scientific Reports, 3*(2522).
51. Weng, L., Flammini, A., Vespignani, A., & Menczer, F. (2012). Competition among memes in a world with limited attention. *Nature Scientific Reports*, 2(335).
52. Kossinets, G., & Watts, D. J. (2009). Origins of homophily in an evolving social network1. *American Journal of Sociology, 115*(2), 405–450.
53. McPherson, J. M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations: Status distance and the composition of face-to-face groups. *American Sociological Review, 52*(3), 370–379.
54. Aral, S., Muchnik, L., & Sundararajan, A. (2009). Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences (PNAS), 106*(51), 21544–21549.
55. Şimşek, O., & Jensen, D. (2008). Navigating networks by using homophily and degree. *Proceedings of the National Academy of Sciences (PNAS), 105*(35), 12758–12762.
56. Crandall, D., Cosley, D., Huttenlocher, D., Kleinberg, J., & Suri, S. (2008). Feedback effects between similarity and social influence in online communities. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)* (pp. 160–168).
57. Jamieson, K. H., & Cappella, J. N. (2009). *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford: Oxford University Press.
58. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Flammini, A., & Menczer, F. (2011). Political polarization on twitter. In *Proceedings of 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
59. Shalizi, C., & Thomas, A. (2011). Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods & Research, 40*(2), 211–239.
60. Granovetter, M. (1995). *Getting a job: A study of contacts and careers*. Chicago: University of Chicago Press.
61. Bakshy, E., Rosenn, I., Marlow, C., & Adamic, L. (2012). The role of social networks in information diffusion. In *Proceedings of ACM International World Wide Web Conference (WWW)* (pp. 519–528).
62. Brown, J., & Reingen, P. (1987). Social ties and word-of-mouth referral behavior. *Journal of Consumer Research, 14*(3), 350–362.
63. Friedkin, N. (1980). A test of structural features of granovetter's strength of weak ties theory. *Social Networks, 2*(4), 411–422.

64. Granovetter, M. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory, 1*(1), 201–233.

65. Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective knowledge transfer. *Management Science, 50*(11), 1477–1490.

66. Lin, N., Ensel, W. M., & Vaughn, J. C. (1981). Social resources and strength of ties: Structural factors in occupational status attainment. *American Sociological Review, 46*, 393–405.

67. Nelson, R. E. (1989). The strength of strong ties: Social networks and intergroup conflict in organizations. *Academy of Management Journal, 32*(2), 377–401.

68. Onnela, J. -P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., et al. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences (PNAS), 104*(18), 7332–7336.

69. Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of ACM International Conference on Human Factors in Computing Systems (CHI)* (pp. 211–220).

70. Wellman, B., & Wortley, S. (1990). Different strokes from different folks: Community ties and social support. *American Journal of Sociology, 96*(3), 558–588.

71. Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E., et al. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature, 489*(7415), 295–298.

72. Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.

73. Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology, 9*(10), 178–190.

74. Huberman, B., Romero, D., & Wu, F. (2009). Social networks that matter: Twitter under the microscope. *First Monday, 14*(1), 8.

75. Gonçalves, B., Perra, N., & Vespignani, A. (2011). Modeling users' activity on Twitter networks: Validation of Dunbar's number. *PLoS One, 6*(8), e22656.

76. Wu, F., & Huberman, B. A. (2007). Novelty and collective attention. *Proceedings of the National Academy of Sciences (PNAS), 104*(45), 17599–17601.

77. Perra, N., Baronchelli, A., Mocanu, D., Gonçalves, B., Pastor-Satorras, R., & Vespignani, A. (2012). Random walks and search in time varying networks. *Physical Review Letters, 109*, 238701.

78. Shaw, L. B., & Schwartz, I. B. (2010). Enhanced vaccine control of epidemics in adaptive networks. *Physical Review E, 81*, 046120.

79. Volz, E., & Meyers, L. A. (2009). Epidemic thresholds in dynamic contact networks. *Journal of the Royal Society Interface, 6*, 233241.

80. Billingsley, P. (1995). *Probability and measure* (p. 362). New York: Wiley.

81. Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: reasoning about a highly connected world*. Cambridge: Cambridge University Press.