How People Interact in Evolving Online Affiliation Networks

Lazaros K. Gallos,[1] Diego Rybski,[1,2] Fredrik Liljeros,[3] Shlomo Havlin,[4] and Hernán A. Makse[1]

[1]*Levich Institute and Physics Department, City College of New York, New York, New York 10031, USA*
[2]*Potsdam Institute for Climate Impact Research, 14469 Potsdam, Germany*
[3]*Department of Sociology, Stockholm University, S-10691, Stockholm, Sweden*
and Institute for Futures Studies - Box 591, SE-101 31 Stockholm, Sweden
[4]*Department of Physics, Bar-Ilan University, Ramat Gan 52900, Israel*

The study of human interactions is of central importance for understanding the behavior of individuals, groups, and societies. Here, we observe the formation and evolution of networks by monitoring the addition of all new links, and we analyze quantitatively the tendencies used to create ties in these evolving online affiliation networks. We show that an accurate estimation of these probabilistic tendencies can be achieved only by following the time evolution of the network. Inferences about the reason for the existence of links using statistical analysis of network snapshots must therefore be made with great caution. Here, we start by characterizing every single link when the tie was established in the network. This information allows us to describe the probabilistic tendencies of tie formation and extract meaningful sociological conclusions. We also find significant differences in behavioral traits in the social tendencies among individuals according to their degree of activity, gender, age, popularity, and other attributes. For instance, in the particular data sets analyzed here, we find that women reciprocate connections 3 times as much as men and that this difference increases with age. Men tend to connect with the most popular people more often than women do, across all ages. On the other hand, triangular tie tendencies are similar, independent of gender, and show an increase with age. These results require further validation in other social settings. Our findings can be useful to build models of realistic social network structures and to discover the underlying laws that govern establishment of ties in evolving social networks.

DOI: 10.1103/PhysRevX.2.031014   Subject Areas: Complex Systems, Interdisciplinary Physics, Statistical Physics

I. INTRODUCTION

Uncovering patterns of human behavior addresses fundamental questions about the structure of the society we live in. The choices made at the individual level determine the emergent complex global network underlying a given social structure [1]. Conversely, the structure of the social network that constitutes an individual's community also affects to a large extent the individual's ability to act. For instance, the position in the network structure may facilitate one's ability to interact with others by providing information of possible choices and their consequences [2], or by supplying the individual with different kinds of material and immaterial resources [3]. On the other side, this structure may also limit this individual's ability to act by excluding information [2] through local social norms and through social control.

Detecting regularities and motifs in the development of social networks provides significant tools for the understanding of the structure of society. For example, the SIENA approach is a widely used tool that comprises a variety of statistical analyses of network data with empha-

sis on actor-oriented models, focusing mainly in small networks of approximately 10 to 1000 nodes [4]. Kossinets and Watts [5] study the ties between university faculty and students. Ties are detected through multiple Email exchanges over a period of 60 days, and the existence of ties is then shown to depend on how many classes are shared, common interests, etc. More recently, Kovanen *et al.* [6] have studied the order by which temporal motifs, such as triangles, squares, etc., are created. Szell and Thurner [7] have studied the interactions in an online game over three years, by creating the networks of in-game friends, enemies, and user communications. In addition to a number of structural features, the authors study the creation of triangles and the time evolution from a given triangle profile to another. A number of statistical association models are also widely used to link a social network structure to a statistically significant social mechanism of interaction [8]. Social theoretical frameworks [9], such as the multitheoretical multilevel (MTML) formalism [10], have proposed a set of mechanisms of social interaction to describe the probabilistic tendencies of creation, maintenance, dissolution, and reconstitution of interpersonal ties during the evolution of a social network. Examples of mechanisms include [see Fig. 1(a)]: (1) reciprocity (named *social exchange* after the most likely social mechanism), (2) friend of a friend ties or closing triangles (*balance*), (3) exploration of distant network areas which require at
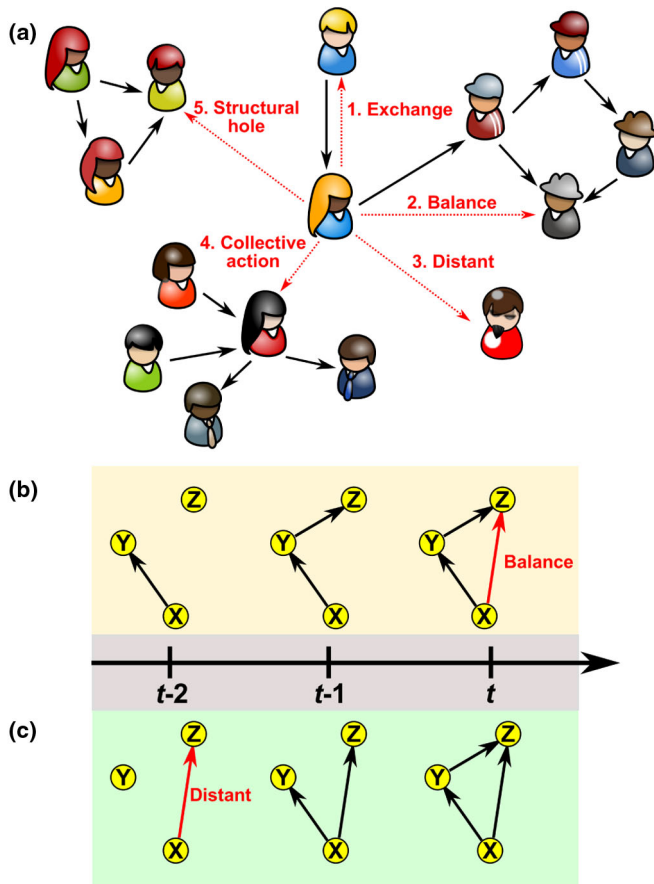
FIG. 1. (a) The five probabilistic tendencies we used to classify the interactions. Black arrows indicate existing links and red arrows are the possible options for a new link, according to the following tendencies: (1) social exchange, which corresponds to establishing a reciprocal link, i.e., add as favorite someone who has already added us to their favorites lists; (2) balance, where we select a favorite who is in the list of one of our existing favorites (friend of a friend); (3) distant connection, where the connection is to a member with whom there is no proximity, i.e., one needs at least three links to reach this member; (4) collective action, where we connect to a person whose connectivity is well above the average connectivity in the community (we quantify this behavior by examining whether the total degree of the receiving agent belongs to the upper 5% of the degree distribution at the given time); and (5) structural hole, where a link connects two otherwise not connected clusters of at least three members each, and which are otherwise not directly linked to each other (in the picture this link would connect the cluster of people in hats with the red-haired cluster). (b),(c) Why we cannot extract tendencies from a static snapshot: In the presented example a triangle relation is built from time $t-2$ to time t under two different scenarios that lead to the same resulting triangle. (b) The ties X-Y and Y-Z can be formed, at times $t-2$ and $t-1$, respectively, via distant mechanisms resulting in a balance mechanism for the formation of X-Z at time t. Here X uses a friend of a friend to be introduced to Z. (c) A different path, though, would classify the X-Z tie differently. If X connects to Z before connecting to Y, then the X-Z link represents a distant tendency, since there are no close connections between them. A static network analysis would suggest that X used balance to connect to Z, instead.

least three steps from the position of the person in the current network (*self-interest* theories), (4) ties facilitating dissemination of information by linking to well-connected people (named *collective action* or *preferential attachment* [11]), and (5) links that act as bridges between two subnetworks that are not directly linked (*structural hole* mechanism). Contractor *et al.* [10] have further identified a set of probabilistic tendencies for ties to be present or absent in networks that the different families of theoretical mechanisms may cause. One important conclusion [10] is that a given family of theoretical mechanisms may generate different probabilistic tendencies for ties to be present or absent. Furthermore, the same probabilistic tendency may be caused by several different families of theoretical mechanisms.

In the present study we aim to unravel significant patterns in these social mechanisms of human interaction by monitoring and analyzing the time evolution of the actions of members of two online affiliation networks. The term affiliation refers to data based on comembership or coparticipation in events, where here members use the Internet to interact with each other through the online sites [12]. A connection in such sites may indicate underlying social ties [13].

In principle, a formal statistical analysis, such as exponential random graph models [8,14] would search for regularities or motifs in the social structure by comparing a static snapshot of the network with a suitable ensemble of equiprobable random configurations. However, this approach cannot characterize the decisions taken (consciously or not) at the individual level on the type of mechanism used for an established connection. A direct application of a statistical analysis to evolving networks may not be able to resolve the full spectrum of human interactions. This is due to the inherent history-dependent nature of social interactions, i.e., the interaction mechanisms determine the evolving network, which, in turn, conditions the human choices of interaction. Figures 1(b) and 1(c) illustrate this point during the generation of a hypothetical triangular XYZ relation at time t. This static pattern may be associated with a balance mechanism for the tie XZ (friend of a friend) as a result of closing the triangle as shown in Fig. 1(b). However, a closer inspection of the time evolution of tie formation reveals the possibility of a different classification of the XZ link, where agent X has used the distant mechanism at time $t-2$ to connect with Z as in Fig. 1(c). The above example can be also understood in a real-world situation. In a social setting, such as trying to date someone or applying for a job, applicants have a different perspective if they are introduced through a common acquaintance or if there is no former connection. Since we eventually may end up with the same balance triangle, there is no way for the analysis of the static pattern to distinguish between the two separate perspectives.

The above example can be generalized to the global network level. For instance, an agent may decide to connect to agents that are far away in the network (distant mechanism). Eventually, individuals are brought closer to each other to form a tightly connected cluster. The evolving nature of the network may change those initial distant interactions into balance, as new relations are created in the network. Therefore, the precise knowledge of the time evolution of each tie in the network is crucial to unravel the relevant behavioral mechanisms in a real community.

Here, we present a microscopic and temporal statistical analysis of the evolution of two online social networks; one from its original inception and the other after it is well developed. We aim to uncover how the combination of different social mechanisms eventually shapes the interaction network. Our longitudinal approach focuses on characterizing each interpersonal tie at the time when it is established. The knowledge of the order in which each link was formed allows us to characterize social patterns that cannot be derived from statistical analysis of static snapshots of the networks.

II. DATA SETS AND METHODS

We study the affiliation networks of two online social networking sites in Sweden, *pussokram.com* [15] and *qx.se* [16]. Both data sets were de-identified in their source. The pussokram community (POK for brevity) is used mainly by Swedish young adults for friendship, including dating and nonromantic relations. Activity in the community was recorded for 512 consecutive days, starting on the day that the site was created in 2001. At the end of recording, the community had 28 876 members with a mean user age of 21 years who have performed ∼190, 000 interactions. The QX site is the Nordic region's largest and most active web community for homosexual, bisexual, and transgender people. The site is also frequently used by heterosexual men and women. Activity among the users was recorded during two months starting in November, 2005. At that time there were about 180 000 registered members; 80 426 of them were active during the recording period establishing more than 1 million ties.

These online services address either adults or teens who intend on meeting new friends. Decisions about connections in the site are based on information provided in the site itself and typically do not reflect preexisting relations in real life. It is unlikely that members would contact real-life friends to any significant extent while seeking romance or dating online. This is in contrast to sites such as Facebook, where the majority of connections are transferred from existing offline acquaintances.

There are many types of interactions between members in the two communities under study, but we focus on those which imply a firmer commitment than, e.g., simply sending a message [17]. Such interactions are (a) the favorites list in QX, and (b) the guest-book signing in QX and POK.

The former interaction represents a clear declaration of approval and/or interest, while the latter is a communication publicly accessible to all community members where a link does not necessarily indicate a particularly close relationship. We compare two means of interaction in one community (favorites list and guest-book signing in QX) and the same type of interaction (guest-book signing) in two communities (QX and POK). We use the guest-book signing to test consistent trends in the results.

In the QX data set, it is possible that a user can remove a contact at any point. There was a small number of such links, in total less than 1% of the total links, that were removed during our monitoring window. It may be interesting to study the conditions of ties removal in parallel with the addition process, but the small number of removed contacts does not influence our results here, and we do not pursue this topic further.

Each individual knows the following structural information from the affiliation network: (a) who has added her in their favorites list or who has written in her guest book, (b) the members that she has added in her favorites list, and (c) the friends of her friends since the user can access the favorites list of friends. This subnetwork defines the immediate neighborhood of a member. Actions involving this neighborhood are captured by social exchange and balance mechanisms. The members situated farthest away from this immediate neighborhood are considered to belong to the rest of the network for which the user has no direct information. Interactions with these members are classified as distant. A collective action can also be a conscious choice, since a member can assess the popularity of others through access to their favorites list, but it is also possible that this action may not be conscious. A structural hole requires a much wider knowledge of the network structure, and thus it is the only mechanism that a member does not realize that she is using.

Our analysis can be readily extended to treat more general situations. For simplicity, though, here we will not evaluate exogenous mechanisms where interactions are based on attributes of the actors, such as homophily, common interests, etc. [10]. We will not study further the effect of focus constraints, i.e., the increased likelihood of a tie being present among people who share a social context, for example, living close to each other geographically or working at the same office [18]. The crux of the matter is to quantify the different probabilistic tendencies about the actions of the users as they are determined by the knowledge of the user about the structure of the affiliation network that is the vital part of her social life in the community.

The detailed quality of our longitudinal data allows us to identify the precise probabilistic tendencies for tie formation that a newly established link corresponds to, when an actor adds a new favorite to her list (or signs a guest book). Every interaction that occurred between two members was recorded together with the timestamp when the event took

place. We create the evolving network of interacting agents by adding the directed links in sequential order. For example, at the time when a member X adds a member Y in the favorites list of X, we create a directional link from X to Y. Similarly, in guest-book signing, the directional link from X to Y corresponds to X writing in Y's guest book (we take into account only the first time X signs Y's guest book and ignore repeated signings). Every time we add a link, we characterize this action according to the probabilistic tendencies described in Fig. 1(a), as dictated by the network configuration at the given moment. Every link is therefore assigned to one or more probabilistic tendencies: exchange, balance, distant, collective action, and structural hole. We define the probabilities of each tendency P_{exc}, P_{bal}, P_{dis}, P_{ca}, and P_{sh} respectively, as the number of links that were created using the corresponding tendency normalized by the total number of links created up to a given time t.

A newly formed link is assigned to the exchange tendency when it is established in the opposite direction of an existing link. The balance tendency corresponds to a directed network distance $\ell = 2$, i.e., when a link points to a friend of a friend (ℓ is the directed distance between two nodes just before the link is formed—defined as the shortest path with all arrows pointing to the same direction, so that a directed path exists between these two nodes). If the distance between the two nodes is $\ell \geq 3$, the link represents the distant tendency. A link is considered as collective action when the chosen node is a hub. We define a hub as a node whose total degree (counting both incoming and outgoing links) belongs to the upper 5% of the degree distribution as measured at the time of link formation. A link represents the structural hole tendency when this link connects two clusters of at least three members that would otherwise be disconnected. Table I summarizes these definitions.

In general, the increase in the probability of a tie forming under a given tendency will not necessarily be compensated for by a tie with decreased probability under another tendency. The relative probabilities between tendencies do not necessarily present competing risks, and different tendencies may act at the same time. It is then possible that one link jointly represents more than one type of tendency in tie formation. In this case, we assign this action to all involved tendencies. For instance, a balance tie could be also catalogued as collective action if the agent closes a triangle by connecting to a hub. Based on the definitions, only balance and distant tendencies are complementary to each other ($P_{bal} + P_{dis} = 1$) so that the presence of one excludes the presence of the other. The other tendencies are normalized as, e.g., $P_{ca} + P_{not-ca} = 1$ (P_{not-ca} is the probability of not performing a collective action).

By establishing all links in the order they appeared, we can recreate the entire history of the directed network of interactions. While POK starts at $t = 0$ from an empty network, QX has a large part of the network already in place at $t = t_0$, our initial recording date. In this case, we know all the existing links at $t = t_0$. Thus, in QX, we characterize only the network links that were added during the monitoring period.

Figure 2 presents the fraction of appearance of each tendency when considering all recorded interactions in the studied data sets, QX and POK, and the means of interaction—guest book and favorites list.

The results are fairly independent of the specific community and the means of interaction. The probabilities P_{exc}, P_{bal}, and P_{ca}, appear each at approximately 15–30% of all actions. The distant mechanism is dominant, with $P_{dis} \approx 80\%$ of the established links. Collective action remains low at $P_{ca} \approx 20\%$ considering that this tendency is considered the main driver in some models of network formation through preferential attachment [11,19]. A very small fraction of links P_{sh} "fills" the structural holes. This is a result of the small numbers of clusters that exist in each community, so that the chances to connect isolated clusters are small. In particular, comparison to the random case (where the same members act at each time step, but instead of the established link they choose a random connection, Fig. 2, yellow bars) reveals that the structural hole tendency is more probable when an agent connects to a random member. In other words, although there exist opportunities for a structural hole, the members tend to stay within their own subnetworks, despite the lack of knowledge on the global structure. The percentages for the other tendencies are also very different from random selections. This implies that community members follow social criteria when adding new

TABLE I. List of tendencies, indicators, and the type of directionality in the network used to detect the tendency. ℓ is the distance between two nodes as measured by the shortest path in the directed network.

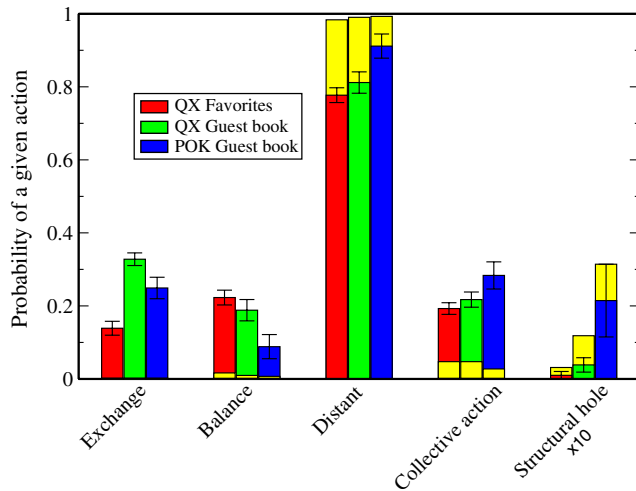|	Tendency	Indicator	Directionality
P_{exc}	Social Exchange	$\ell = 1$, mutual link	Directed
P_{bal}	Balance	$\ell = 2$	Directed
P_{dis}	Distant	$\ell \geq 3$	Directed
P_{ca}	Collective Action	Link to a hub	Undirected
P_{sh}	Structural Hole	Connect two clusters	Undirected

FIG. 2. The relative appearance of the five probabilistic tendencies in the actions of the community members in QX using favorites (red), in QX using guest book (green), and in POK using guest book (blue). These tendencies are compared to a completely random selection (yellow). Exchange and balance are practically nonexistent in random selections, but carry significant weight in the interactions of the real communities. Connecting to distant members appears in the community much less frequently than in random, while the preference towards well-connected agents (collective action) is significantly more prominent. Finally, the structural hole is significantly suppressed in the real communities compared to the randomized case.

favorite members (or sign guest books). We verified the robustness of our results by comparing the percentages of the links at the early stages of network formation with those of the links that were established later in the process. For example, in QX favorites the first half of the actions data set gives practically the same result as the second half: exchange was 13.8% for the first half and 13.9% for the second, balance was used 22.1% versus 22.4%, and collective action was used 18.8% versus 19.7%. Furthermore, the stability of this result over the evolution of the links is verified later, in Sec. III E.

Our analysis has shown that the direct calculation of the tendencies of link formation from the time evolution of the network provides a consistent characterization of the social mechanisms involved, which is different from a static snapshot. Furthermore, the present analysis allows us to determine if the found tendencies are influenced by important actor attributes that are hypothesized to have an association with ties formation [20]. These attributes include age, gender, popularity, and activity intensity measured as the number of links developed at a given time. Next, we incorporate these attributes in our analysis to attempt to understand how different factors influence the behavior of the actors. We show that the gender, age, activity intensity, and popularity can lead to a different probability of using a given tendency.

III. RESULTS

A. Gender influence

Our analysis reveals that gender is an important attribute determining the social tendencies. Analysis of the QX community (the only one reporting gender) reveals that men do not use some mechanisms in the same way as women (Fig. 3). Using the gender information in the QX favorites lists, we find that a female member is almost 3 times more likely to have an exchange tendency compared to male members and 3 times more probable to fill structural holes (men, on the other side, perform distant and collective actions at higher percentages). The significant difference in exchange, for example, reveals a different approach of online communication between men and women [21]. Our result is in agreement with the self-reported tendency of women users to exchange more private Emails than participate in public discussions [22]. The stronger preference for exchange of female users in the QX community can also be seen as a similar trait where women tend to develop stronger interpersonal relations by frequently reciprocating friendships.

B. Age influence

In the databases that we studied, members of different ages tend to present different behaviors. In Fig. 4 we calculate the fraction of actions that correspond to a tendency as a function of the self-reported age of the QX members. In the insets, we separate the corresponding probabilities for male and female members. We find that while reciprocity in women remains high as they age, men instead reduce it by a factor of 2 as they reach 40 years of age. This shows that younger male members are more eager to reciprocate their connections. In contrast, the level of balance is roughly constant for both genders and independent of age, with an
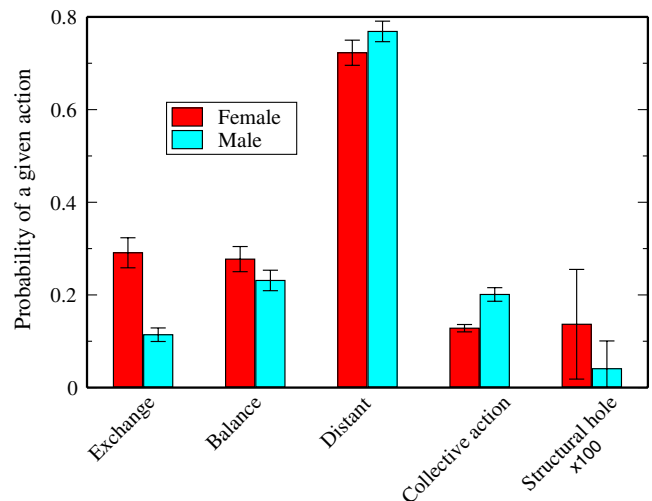

FIG. 3. Probability of different tendencies, based on self-reported gender in the QX community in favorites list interactions. Exchange and structural hole are significantly more frequent in females compared to males.
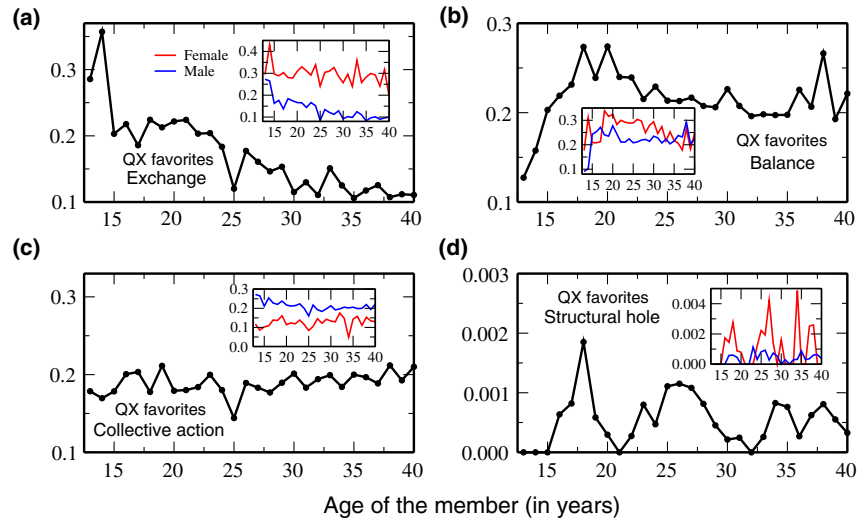
FIG. 4. Variation of the average tendency percentage with the self-reported age of the QX members. (a) The exchange tendency decreases with age. (b) The balance tendency is sharply increasing with age in younger ages, and slowly declines for ages above 20 years. We do not observe any strong dependence on age for (c) collective action or (d) structural hole (bottom right). The insets show the differences between males and females of the same age for each tendency.

important exception at the youngest ages, where members younger than 20 years old are using systematically less balance links. This could be because it is more difficult for them to develop a stable local network in an adult-oriented community. There are no significant trends with age for collective action or structural hole, although the latter tendency is rarely used. The gender-based trends shown in Fig. 3 are consistent with the age-based results. Women of a given age are always using more exchange and less collective action tendencies than men of the same age (insets of Fig. 4).

C. Activity influence

Communities include members of varying activity [17]. In order to study the effect of the different activity levels, we address the question of whether a higher involvement in a community is accompanied by a different pattern in the probabilistic tendencies of social mechanisms. We calculate the different probabilities of social mechanisms as a function of the number of k_{out} outgoing links for each member. For instance, $P_\alpha(k_{out})$ (where α denotes exchange, balance, etc.) measures the probability that the next action will correspond to α, when the member has k_{out} outgoing links. We measure $P_\alpha(k_{out})$ through all the actions of members when they increase the number of outgoing links from k_{out} to $k_{out} + 1$, irrespective of the time that the action was performed. Interestingly, we find that a member typically modifies her behavior according to her current degree of activity k_{out}. As a member becomes more involved in the community and, as a consequence, increases the size of her favorites list or signs more guest books, the member switches to a different relative percentage of using each tendency.

We identify the following pattern which is very consistent across the two data sets and different types of

interactions (see Fig. 5). The first tie of a new member is always distant since the member has no network established. However, even at this stage, 20–30% of these links are also exchange—meaning that a new member readily "responds" to the incoming link by established members—and collective action—meaning that the member immediately searches for popular members in the community. At this earlier stage, balance tendency is suppressed, since linking to friends of friends requires first a firm establishment of the immediate neighborhood.

An interesting crossover appears when the members arrive to a size $k_{out} \approx 10$ in their favorites list (see, for example, Fig. 5(a) for QX favorites). The percentage of all tendencies up to that value is approximately constant. At around 10 interactions in QX favorites, balance overtakes both exchange and collective action in the behavioral tendencies. As the members keep adding more links, the distant mechanism drops significantly to approximately 60% after $k_{out} \approx 100$, and the balance tendency grows increasingly stronger consequently. Similarly, the exchange tendency declines steadily towards 0 as the size of the favorites list increases towards the hundreds. Collective action leading to preferential attachment seems to be the most stable over a longer k_{out}-range. Finally, the relative probability of $P_{sh}(k_{out})$ peaks at low and large values of k_{out}. The structural holes are filled mainly by either new members or well-established members, with a significantly smaller fraction of structural holes performed in the intermediate k_{out} regime. This interesting behavior reveals trends in the social tendencies across the individual users as they enter the network.

The choice of different tendencies is, thus, shown to have a complex dependence on the individual's level of activity. In addition to external attributes, such as gender
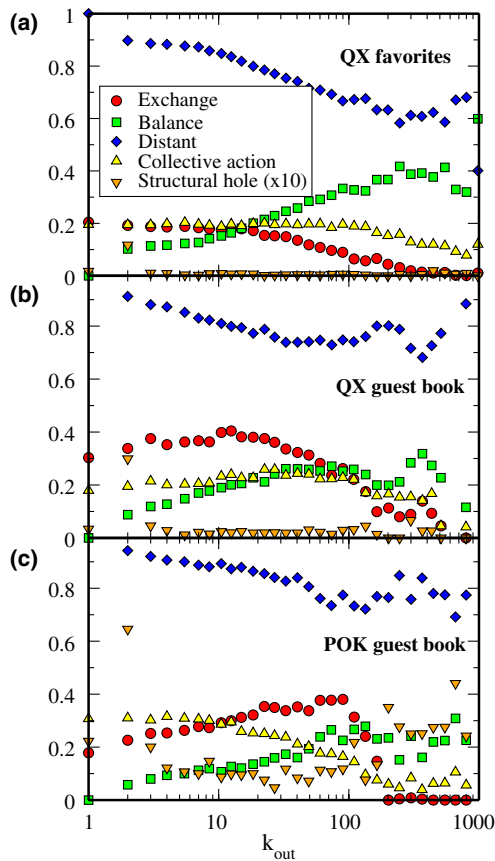
FIG. 5. Fraction of the appearance of a tendency as a function of the adding member's list size, at the time of addition. Qualitatively, all three data sets are in agreement with each other. The small quantitative differences may be due to the different means of interaction and/or the design of each platform.

and age, we find that very active members have different tendencies than the less active ones. Such features can only be extracted by following the entire time evolution of each member's connections.

D. Popularity attributes

So far, our analysis focused on quantifying the different probabilistic tendencies as seen from the member that establishes a link. We characterized the outgoing links which can be controlled by their initiator, in the sense that any member can choose where, when, and how often she connects to other members. However, the popularity (or attractiveness) of a member cannot be adjusted at will. We characterize the popularity based on the number of incoming links. Using the same methodology as above, we can now study how different tendencies determine the popularity of a member.

For each relationship between two people we assign the initiator, i.e., the member who contacted the other member first, and the receiver, i.e., the member who was contacted. In the case of a reciprocal relation we only characterize the link that was established first. Given the list of a member's connections, we can then know what fraction of those

connections is due to the initiative of this member and what fraction originated from the other side. Thus, if someone very often reciprocates but seldom initiates links, she will have a small value of initiated links although she may have a large number of incoming and/or outgoing links.

In Fig. 6(a) we present the histogram of how many members fall into each category. The diagram is roughly divided into three areas: (a) members who initiate a lot of connections but are first contacted by very few members ("spammers"), (b) members who on average equally initiate and receive contacts, and (c) members who receive many more contacts than they initiate ("popular").

The importance of using the time evolution of probabilistic tendencies to determine behavior is reflected in this popularity classification. In Figs. 6(b)–6(d) we present the average percentage for each category and for each tendency that the members use when they add friends themselves. The exchange tendency shows a clear variation with respect to this classification. The popular members in the upper diagonal part of the distribution use a lot of exchange, which can be understood since they respond to friendship requests but rarely start new connections. As we move towards the spammers the exchange tendencies almost disappear, since very few people approach those members and therefore they have a small chance to use exchange. On the contrary, the spammers tend to use balance more, i.e., they connect to friends of friends, since they try to access the largest possible number of the accessible members [Fig. 6(c)]. Finally, connecting to distant parts of the network [Fig. 6(d)] has a more uniform behavior, although the popular members seem to use it more, pointing to a "rich-club" phenomenon [23].

The above-described trends demonstrate the richness of information that becomes accessible by following the evolution of link formation. Nevertheless, we next show that even in the absence of the network history, we can still deduce some useful conclusions on the probabilistic tendencies.

E. Neighborhood landscape change

As discussed above, the presented analysis would not be possible without continuously monitoring the time evolution of the links. The characteristics of a given link with time do not remain necessarily the same as when the connection was established, but they can change due to the addition of more links or the removal of existing ones. For example, a friendship that starts between two isolated individuals may evolve into a densely connected neighborhood, so that a link that began as distant may eventually switch with time to either balance, exchange, collective action, structural hole, or any combination of them.

In order to study how significant the evolution of the link formation tendencies is, we compare the probabilistic tendencies obtained above following the time evolution with those obtained by a statistical analysis of a snapshot of the network. There are a few possible ways to extract
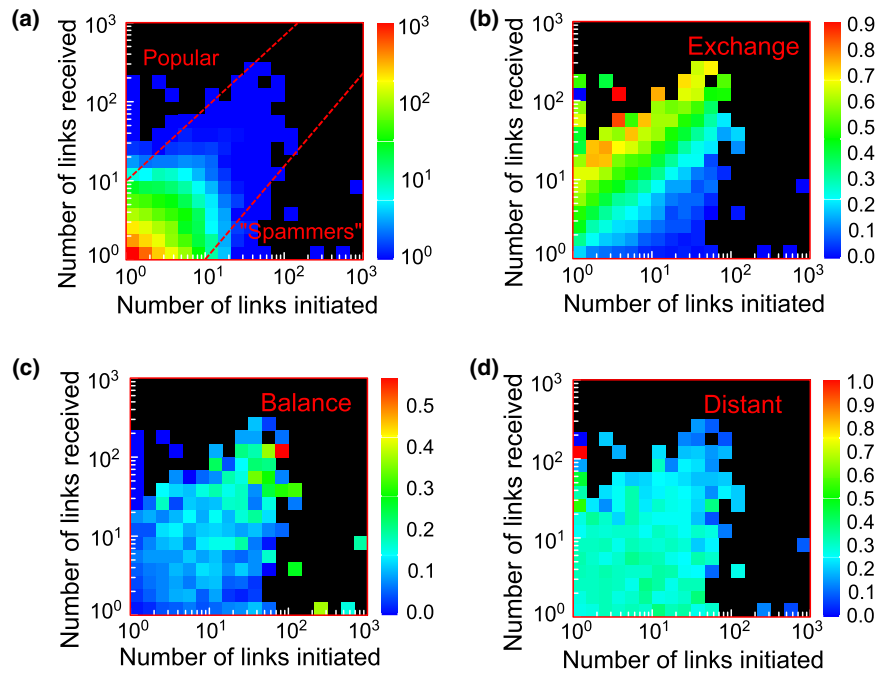
FIG. 6. (a) Histogram of the number of members as a function of the links that they initiated (x-axis) and the links that were pointed to them but initiated at the partner's side (y-axis). (b–d) Average percentage of exchange, balance, and distant mechanisms as a function of the links initiated and received.

information from the static network in order to establish a null model for comparison with the dynamic analysis. For example, one could rebuild a network with completely random choices (Fig. 2), or reshuffle only the order of links addition, or reshuffle the links of one person at a time, or attempt to characterize one link at a time, etc. Each of these processes represents a different perspective and may lead to varying estimation of the static probabilities. As described above, we are here interested in the evolution of the mechanism in any given link, so we characterize all individual links in the static network one-by-one.

The statistical analysis of the static snapshot is done by characterizing all existing links at the given time without using the information from the time when the link was established. Thus, each link is assigned to the specific probabilistic tendencies according to the current neighborhood environment of each agent, independent of the time it was established. We repeat this process for all links in the static snapshot, and we calculate the relative percentage for each mechanism. This is a typical procedure in the literature [24], where analyses of human interactions do not take into account the network evolution and are based instead on a network snapshot. For instance, to estimate the balance mechanism, a standard analysis consists of measuring the number of triangles or balance relations in a static network [24] through the transitivity measure. The percentage of transitive links is estimated through the static network, see, e.g., Ref. [25], where the percentage of transitive links was estimated to be 14.2%. Like a physical system with path-dependent interactions between particles,

the information of the static snapshot of the system is not sufficient to provide the correct statistics of the actions. If the network evolution were known and the number of balance relationships could be calculated exactly as done in our analysis, it would be possible to determine how many of the triangles were really closed in a "balanced" way [according to the example of Fig. 1(b)].

In Fig. 7 we compare the running percentages for each tendency at the moment of addition, such as those
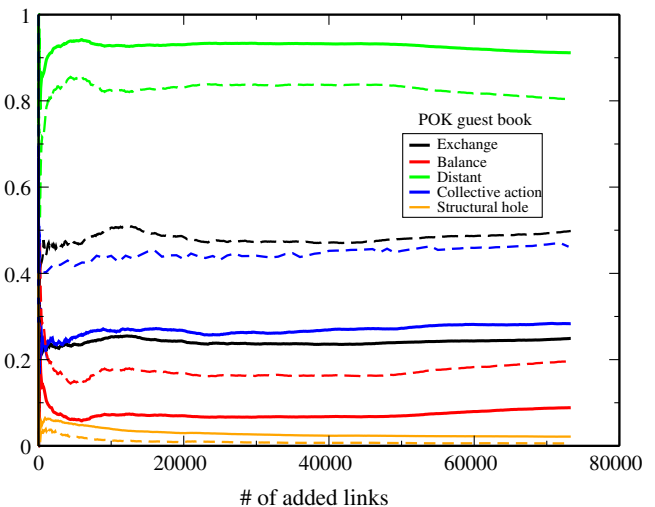

FIG. 7. Comparison of the probabilistic tendencies fraction, where links are characterized either at the time of addition (solid lines) or at the time of observation (dashed lines) in the POK community.

measured in Fig. 2, to those of the corresponding static network. All tendencies are different in these two measurements. Exchange is the only predictable tendency, since by definition it appears 2 times more at the time of observation compared to the time of addition. The other tendencies cannot be predicted from the static measurements. For example, although a member is typically using the balance tendency to add links at a percentage of around 10%, if she tries to evaluate her neighborhood at any point in time she will find out that now approximately 20% of her acquaintances fall under the balance theory. Similarly, the central hubs seem to be re-enforced, since collective action is used in less than 30% of the total actions, but eventually more than 45% of the links are directed towards the biggest hubs. In other words, members are ultimately attached to hubs more often than we could conclude from characterizing their original actions only, due to the dynamic environment. This quantifies and generalizes the situation depicted in Figs. 1(b) and 1(c): the knowledge of the network structure at a given time is not sufficient for characterizing the probabilistic tendencies.

Another aspect of this plot (Fig. 7) is that the tendencies at the time of addition reach their asymptotic values quite fast and they remain roughly constant with time. The corresponding values extracted from the static networks are also quite robust and follow closely the variations of the values in the evolving networks, creating a constant gap between the two curves. Since there is currently no method to estimate the magnitude of the difference between the two cases by static information only, it is still not possible to extract the percentage of the probabilistic tendencies without following the network evolution.

Next, we compare our results with other directed social interaction networks from the literature, such as the Epinions [26], SlashDot [27] and LiveJournal [28] communities. The data sets were downloaded from http://snap.stanford.edu/data. The Epinions data set is a directed network of trust from epinions.com, where a user can declare her trust toward another user, based on submitted reviews. This trust creates a directed link between the two users. The network has 75 879 nodes and 508 837 links. Slashdot.com is a technology-oriented news site, where users can tag each other as friends or foes. In our analysis we only use the friendship links. We use two snapshots of the network, on November 6, 2008 (77 360 nodes and 905 468 links) and on February 2, 2009 (82 168 nodes and 948 464 links) [27]. Finally, Livejournal.com is a social networking site, where users can declare who they consider as their friends. The network that we use has 4 847 571 nodes and 68 993 773 links. For these networks we only have the static snapshots. Therefore, we can only study the exchange tendency, which is the only one that remains unmodified in a static network. (We can always measure the existence of reciprocity, independent of the time it was established.)
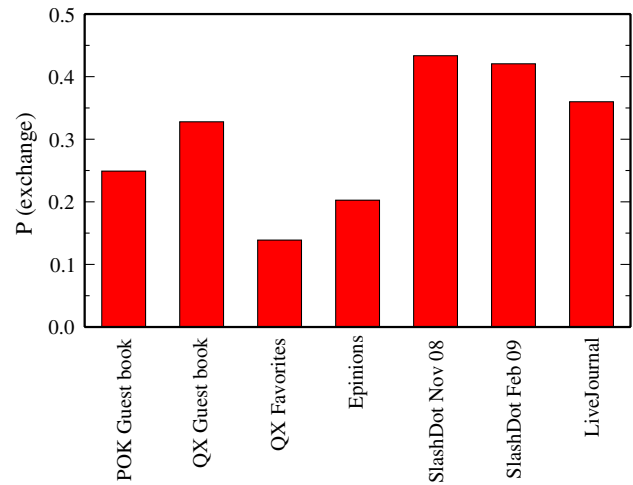

FIG. 8.   Probabilistic exchange tendencies extracted from static network snapshots for several directed networks.

The probability of using the exchange tendency among the different social networks (Fig. 8) depends on the specific features of each community. For example, in the SlashDot and in the LiveJournal communities, where a link shows that a user declares another user as being her friend, there is a large degree of the exchange tendency because mutual relations are favored in these social networking environments. In contrast, in the QX database, the exchange tendency is quite smaller due to the nature of this community. Similarly, in the Epinions database, a link shows that a member trusts the tech reviews of the other member, but this relation is usually not mutual (e.g., if I trust the reviews of an expert reviewer, this reviewer may not necessarily trust my reviews).

IV. DISCUSSION

The wealth of information obtained by our longitudinal analysis can complement other statistical analyses for probabilistic tendencies [10,14]. The family of exponential random graph models [29] (p^*), and, in particular, the logit p^* models [8], have been very successful in analyzing network snapshots at a given moment in time. These methods detect network patterns that appear more frequently than a random null hypothesis would assume. In this way, the underlying mechanisms of network creation are inferred from the resulting motifs. Our present analysis goes beyond this approach by directly facing a number of key issues: we can follow the entire network evolution, we can characterize individual actions, and we can also assign known mechanisms to any given action. The results of these actions often yield network patterns where an individual contribution may be lost in the static snapshot pattern, due to the effect of subsequent connections. In broad terms, our analysis compared to exponential random graph models may be considered to be the analogue of a microscopic statistical physics description compared to a macroscopic thermodynamic approach.

Here, we have shown that following the order of links establishment at the microscopic level in a social network provides a direct measurement of the probabilistic tendencies. This allows both the quantification of the relative strength between tendencies in a given community, and the extraction of useful sociological conclusions. For example, in the communities that we studied, we show that women tend to use the exchange mechanism more frequently than men. This tendency is more pronounced with age since reciprocity in older men largely declines while in women it remains stable across all ages. In these communities, also, men tend to connect to the hubs more often than women, independently of age. The use of triadic closures is almost constant for both genders and all ages, except for the youngest members with ages below 20 years. This may be a consequence of the more adult-oriented character of the community. Similarly, we capture a different use of the tendencies between the more active and less active members. The results that we found characterize the behavior of members in the specific communities studied. A generalization of these results to establish generic trends based on gender, age, etc. would require an extensive study of a large number of social networking sites under different settings. The basis of our findings is that these results cannot be derived analyzing a snapshot of a static network. As shown in Figs. 1(b) and 1(c) and quantified in the preceding section, it is not possible to make assumptions of why a link exists a long time after the link was established.

Our findings reflect the behavior of users in the online networking sites that we studied. The suggested method of following the dynamic evolution, though, represents a consistent method which can be applied to other networks. Further studies in different online communities should elaborate on whether the trends reported here with respect to sex, age, etc., are generic to other types of networks.

The present analysis complements other approaches in the literature [30] by focusing on individual actions and the study of how the underlying mechanisms behind these actions are driving the evolution of the large-scale social network. The ability to isolate individual actions can be also very useful in studying behaviors that are unusual, and help characterize idiosyncratic ways of building the friendship network. The present analysis can be extended to exogenous mechanisms, as well, by incorporating information from other aspects of the activity in the community (e.g., joining specific clubs, participating in forum discussions, communities, etc.). Moreover, our approach can find applications to many complex systems, where links are evolving over time, such as the structural evolution of the Internet, the World Wide Web, etc. For example, this methodology can be used to study the evolution of protein-interaction networks in biology, where one is interested in estimating the different interaction motifs [31].

[1] T. S. Schelling, *Micromotives and Macrobehavior* (Norton, New York, 1978).

[2] J. S. Coleman, E. Katz, and H. Menzel, *The Diffusion of a New Drug among Physicians*, Sociometry **20**, 253 (1957).

[3] P. Bourdieu, *What Makes a Social Class? On the Theoretical and Practical Existence of Groups*, Berk. J. Sociol. **32**, 1 (1987) [http://www.jstor.org/stable/41035356].

[4] T. A. B. Snijders, G. G. van de Bunt, and C. E. G. Steglich, *Introduction to Stochastic Actor-Based Models for Network Dynamics*, Soc. Netw. **32**, 44 (2010).

[5] G. Kossinets and D. J. Watts, *Empirical Analysis of an Evolving Social Network*, Science **311**, 88 (2006).

[6] L. Kovanen, M. Karsai, K. Kaski, J. Kertesz, and J. Saramaki, *Temporal Motifs in Time-Dependent Networks*, J. Stat. Mech. (2011) P11005.

[7] M. Szell and S. Thurner, *Measuring Social Dynamics in a Massive Multiplayer Online Game*, Soc. Netw. **32**, 313 (2010).

[8] S. Wasserman and P. Pattison, *Logit Models and Logistic Regressions for Social Networks. I: An Introduction to Markov Graphs and p^**, Psychometrika **60**, 401 (1996).

[9] P. R. Monge and N. S. Contractor, *Theories of Communication Networks* (Oxford University Press, New York, 2003).

[10] N. S. Contractor, S. Wasserman, and K. Faust, *Testing Multitheoretical, Multilevel Hypotheses about Organizational Networks: An Analytic Framework and Empirical Example*, Acad. Manag. Rev. **31**, 681 (2006).

[11] A-L. Barabasi and R. Albert, *Emergence of Scaling in Random Networks*, Science **286**, 509 (1999).

[12] S. P. Borgatti and D. Halgin, in *The Sage Handbook of Social Network Analysis*, edited by P. Carrington and J. Scott (Sage Publications, Thousand Oaks, CA, 1996).

[13] A. Davis, B. Gardner, and R. Gardner, *Deep South: A Social Anthropological Study of Caste and Class* (University of Chicago Press, Chicago, 1941).

[14] G. Robins, T. Snijders, P. Wang, M. Handcock, and P. Pattison, *Recent Developments in Exponential Random Graph (p^*) Models for Social Networks*, Soc. Netw. **29**, 192 (2007).

[15] P. Holme, C. R. Edling, and F. Liljeros, *Structure and Time Evolution of an Internet Dating Community*, Soc. Netw. **26**, 155 (2004).

[16] X. Lu, L. Bengtsson, T. Britton, M. Camitz, B. J. Kim, A. Thorson, and F. Liljeros, *The Sensitivity of Respondent-Driven Sampling Method*, Journal of the Royal Statistical Society Series A (General) **175**, 191 (2012).

[17] D. Rybski, S. Buldyrev, S. Havlin, F. Liljeros, and H. A. Makse, *Scaling Laws of Human Interaction Activity*, Proc. Natl. Acad. Sci. U.S.A. **106**, 12640 (2009).

[18] S. L. Feld, *The Focused Organization of Social Ties*, Am. J. Sociol. **86**, 1015 (1981).

[19] B. F. De Blasio, A. Svensson, and F. Liljeros, *Preferential Attachment in Sexual Networks*, Proc. Natl. Acad. Sci. U.S.A. **104**, 10762 (2007).

[20] J. G. Parker and S. R. Asher, *Friendship and Friendship Quality in Middle Childhood: Link with Peer Group Acceptance and Feelings of Loneliness and Social Dissatisfaction*, Dev. Psychol. **29**, 611 (1993).

[21] S. C. Herring, *Gender Differences in CMC: Findings and Implications*, Computer Professionals for Social Responsibility Journal **18** (2000) [http://www.cpsr.org/issues/womenintech/herring/].

[22] D. L. Hoffman, W. D. Kalsbeek, and T. P. Novak, *Internet and Web Use in the U.S.*, Commun. ACM **39**, 36 (1996).

[23] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, *Detecting Rich-Club Ordering in Complex Networks*, Nature Phys. **2**, 110 (2006).

[24] S. Wasserman and K. Faust, *Social Network Analysis* (Cambridge University Press, Cambridge, 1994).

[25] N. A. Christakis and J. H. Fowler, *Social Network Sensors for Early Detection of Contagious Outbreaks*, PLoS ONE **5**, e12948 (2010).

[26] M. Richardson, R. Agrawal, and P. Domingos in *Proceedings: The Semantic Web: ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, 2003, Proceedings*, Lecture Notes in Computer Science No. 2870, edited by D. Fensel, K. P. Sycara, and J. Mylopoulos (Springer, New York, 2003), p. 351.

[27] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, *Community Structure in Large Networks: Natural Cluster Sizes and the Absence of Large Well-Defined Clusters*, Internet Math. **6**, 29 (2009).

[28] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan, in *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '06* (ACM, New York, 2006), p. 44.

[29] O. Frank and D. Strauss, *Markov Graphs*, J. Am. Stat. Assoc. **81**, 832 (1986).

[30] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins, in *KDD '08: Proceedings of The 14th ACM SIGKDD International Conference of Knowledge Discovery and Data Mining* (ACM, New York, 2008), p. 462.

[31] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, *Network Motifs: Simple Building Blocks of Complex Networks*, Science **298**, 824 (2002).