RESEARCH ARTICLE

Competition between Homophily and Information Entropy Maximization in Social Networks

Jichang Zhao[1]*, **Xiao Liang**[2], **Ke Xu**[3]

1 School of Economics and Management, Beihang University, Beijing, China, **2** Key Laboratory of Technology in Geo-spatial Information Processing and Application System, Institute of Electronics, Chinese Academy of Sciences, Beijing, China, **3** State Key Lab of Software Development Environment, Beihang University, Beijing, China

* jichang@buaa.edu.cn

Abstract

In social networks, it is conventionally thought that two individuals with more overlapped friends tend to establish a new friendship, which could be stated as homophily breeding new connections. While the recent hypothesis of maximum information entropy is presented as the possible origin of effective navigation in small-world networks. We find there exists a competition between information entropy maximization and homophily in local structure through both theoretical and experimental analysis. This competition suggests that a newly built relationship between two individuals with more common friends would lead to less information entropy gain for them. We demonstrate that in the evolution of the social network, both of the two assumptions coexist. The rule of maximum information entropy produces weak ties in the network, while the law of homophily makes the network highly clustered locally and the individuals would obtain strong and trust ties. A toy model is also presented to demonstrate the competition and evaluate the roles of different rules in the evolution of real networks. Our findings could shed light on the social network modeling from a new perspective.

Introduction

The last decade has witnessed tremendous research interests in complex networks [1–3], including the evolution of social networks [4–8]. It has been found that in many social networks from different circumstances, the probability of having a friend at a distance r is $p(r) \propto r^{-1}$, which is stated as the spacial scaling law [9]. Recent work [10] presents a possible origin that explains the emergence of this scaling law with the hypothesis of maximum information entropy with energy constrains. The authors assume that human strategic behavior is based on gathering maximum information through various activities and being an essential component of the human social behavior, making friends is intuitively one of its significant pathways. However, it is also found conventionally that homophily leads to connections in social

networks [5, 6, 11–17]. Homophily is the principle that a contact between similar individuals occurs at a higher rate than among dissimilar ones [6]. For instance, in social networks, two individuals with more common friends are easier to get connected, where the number of overlapped friends could represent the strength of homophily [17, 18]. Both of the above rules might drive the growth of the network in local structure simultaneously, however, to our best knowledge, little has been done to unveil the relationship between them. We argue that understanding the interplay between these two rules could help reveal the generation of different social ties and shed light on modeling social networks from a new perspective. Therefore, in this paper, we try to fill this gap from the perspective of network evolution in local structure.

Results

Theoretical Analysis

A social network can be modeled as a simple undirected graph $G(V, E)$, where V is the set of individuals (nodes) and E is the set of friendships (ties) among them. As shown in Fig 1a, node 1 may obtain information from nodes 2, 3, 4 and their friends 5, 7. Therefore, as defined in [10], the information sequence for node 1 is {2, 3, 4, 5, 7} and the frequency of each node appears in the sequence is $q_2 = q_3 = q_4 = q_5 = q_7 = 1/5$ for nodes 2, 3, 4, 5 and 7 respectively, while $q_6 = 0$ for node 6. Then the information entropy for node 1 can be obtained as

$$\epsilon(1) = -\sum_{i=1}^{7} q_i \log q_i = 1.61.$$

Next, we assume the social network evolves to the one as shown in Fig 1b under the rule of homophily. For example, node 1 and node 5 may establish a new friendship because they share the common friend node 2. Therefore, the updated information sequence for node 1 is {2, 3, 4, 5, 5, 7, 2} currently. Then the new frequency of each node appears in the sequence is $q_2' = q_5' = 2/7, q_3' = q_4' = q_7' = 1/7$, and $q_6' = 0$. We recompute the information entropy of node 1 as depicted above and obtain

$$\epsilon^{'}(1) = -\sum_{1}^{7} q_i{}^{'} \log q_i{}^{'} = 1.55.$$

It can be easily observed that $\Delta\epsilon(1) = \epsilon^{'}(1) - \epsilon(1) < 0$ after node 1 built a new tie with node 5, which means in the evolution dominated by homophily, the information entropy for node 1 decreases. It is an intuitive observation that the rule of homophily is incompatible with the law of maximum information entropy, and a general explanation is introduced as follows. Note that here we mainly discuss the network evolution in local structure, in which ties are newly built only with nodes two hops away. Because of this, with the aim of simplification, conditions of limited energy and nodes' distances are not considered in the following analytical framework. Besides, the magnificent development of the online social network has facilitated our daily social activity greatly [19, 20], so here the cost of establishing a new tie is assumed to be a constant and it is independent to the distance in social networks.

We define $n(i)$ as the set of individual i's initial friends and k_i is i's degree, i.e., the number of its friends. Then the set of overlapped friends between i and j is $c(i, j) = n(i) \cap n(j)$ and $c_{ij} = |c(i, j)|$ is the number of their common friends. We define $U = \cup_{q \in n(i)} n(q) \cup n(i)$. We also define $\Psi = \{j\} \cup c(i, j)$, where j is a random individual appearing in i's information sequence s (i) and $j \notin n(i)$. Based on the definition of information entropy in [10], we can obtain the
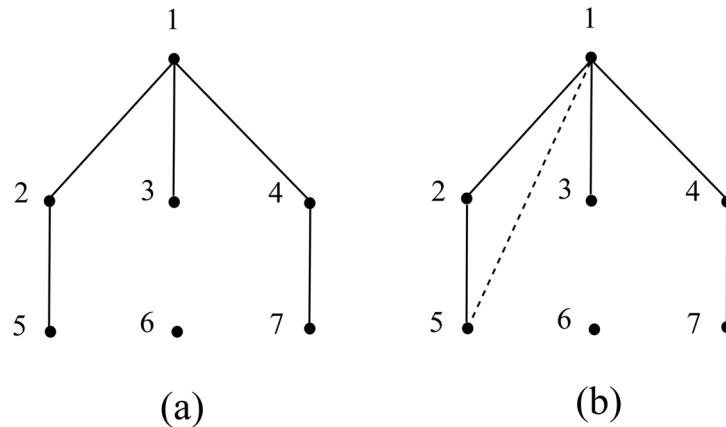
Fig 1. A simple example of the network evolution driven by homophily in local structure.

information entropy for node i is

$$\epsilon(i)_j = \quad - \quad \sum_{q \in U/\Psi} \frac{n_q}{s_i} \log \frac{n_q}{s_i} - \sum_{l \in c(i,j)} \frac{n_l}{s_i} \log \frac{n_l}{s_i}$$
$$- \quad \frac{c_{ij}}{s_i} \log \frac{c_{ij}}{s_i}, \tag{1}$$

where n_q is the count that q appears in $s(i)$ and s_i is the length of $s(i)$. Since we mainly investigate the evolution in local structure, here only friends of i and friends of its friends are considered during the computation of the entropy. Then we assume that a new friendship is established between i and j and the current entropy for i is

$$\epsilon'(i)_j = \quad - \quad \sum_{q \in U/\Psi} \frac{n_q}{s_i'} \log \frac{n_q}{s_i'} - \sum_{l \in c(i,j)} \frac{n_l + 1}{s_i'} \log \frac{n_l + 1}{s_i'}$$
$$- \quad \frac{c_{ij} + 1}{s_i'} \log \frac{c_{ij} + 1}{s_i'} - (k_j - c_{ij}) \frac{1}{s_i'} \log \frac{1}{s_i'}, \tag{2}$$

where $s_i' = s_i + k_j - c_{ij} + 1 + c_{ij} = s_i + k_j + 1$, which is the length of the updated information sequence, where k_j is the initial degree of j. Therefore, the change of entropy for i caused by the new tie with j, i.e., $\Delta\epsilon(i)_j = \epsilon'(i)_j - \epsilon(i)_j$ could be rewritten as

$$\Delta\epsilon(i)_j = \sum_{q \in U/\Psi} \left(\frac{n_q}{s_i} \log \frac{n_q}{s_i} - \frac{n_q}{s_i'} \log \frac{n_q}{s_i'} \right)$$
$$+ \sum_{l \in c(i,j)} \left(\frac{n_l}{s_i} \log \frac{n_l}{s_i} - \frac{n_l + 1}{s_i'} \log \frac{n_l + 1}{s_i'} \right)$$
$$+ \left(\frac{c_{ij}}{s_i} \log \frac{c_{ij}}{s_i} - \frac{c_{ij} + 1}{s_i'} \log \frac{c_{ij} + 1}{s_i'} \right)$$
$$- (k_j - c_{ij}) \frac{1}{s_i'} \log \frac{1}{s_i'}. \tag{3}$$

Assume $f(x) = x \log x$,

$$f(x + \Delta x) = f(x) + f'(x)\Delta x + o((\Delta x)^2),$$

therefore,

$$\frac{n_l + 1}{s_i'} \log \frac{n_l + 1}{s_i'} = \frac{n_l}{s_i'} \log \frac{n_l}{s_i'} + \left(\log \frac{n_l}{s_i'} + 1 \right) \frac{1}{s_i'} + o\left(\frac{1}{s_i'^2} \right)$$

and

$$\frac{c_{ij} + 1}{s_i'} \log \frac{c_{ij} + 1}{s_i'} = \frac{c_{ij}}{s_i'} \log \frac{c_{ij}}{s_i'} + \left(\log \frac{c_{ij}}{s_i'} + 1 \right) \frac{1}{s_i'} + o\left(\frac{1}{s_i'^2} \right).$$

Then for Eq (3) we have (for details, see S1 Equation),

$$\begin{aligned}
\Delta\epsilon(i)_j = \quad &- \quad \frac{k_j + 1}{s_i'} \epsilon(i)_j - \sum_{l \in \Psi} \frac{1}{s_i'} \log n_l - \frac{c_{ij} + 1}{s_i'} \\
&+ \quad \frac{k_j + 1}{s_i'} \log s_i' - (c_{ij} + 1)o\left(\frac{1}{s_i'^2} \right).
\end{aligned} \tag{4}$$

Suppose that k_j is fixed, it can be easily obtained that as c_{ij} grows, $\Delta\epsilon(i)_j$ decreases. Given the network is undirected, so this conclusion is also proper for j. Then we can conclude that if we build a new tie between i and j, the information entropy gain $\Delta\epsilon(i, j) = \Delta\epsilon(i)_j + \Delta\epsilon(j)_i$ produced by this new friendship for the two nodes decreases as c_{ij} increases. It tells us that for the nodes with more common friends, establishing a new tie between them produces less information entropy gain for them. Be brief, there is a competition between homophily and information entropy in breeding a new connection. Note that $\Delta\epsilon(i, j)$ declining with c_{ij} might be very slow, because generally s_i' is much greater than c_{ij}.

In fact, the information entropy for i represents the diversity of its information sources. If we create ties between i and other nodes who have overlapped friends with it, these nodes will appear more frequently in its information sequence and even become the dominating sources of the information. Then the diversity of the information source is weaken and the gain of the information entropy decays accordingly.

Empirical Analysis

In order to validate the above analysis, we employ several data sets, including both synthetic and real-world networks, for further empirical study. The synthetic data sets are generated by BA [21], Small World [22] and CNNR [14] models. BA is a classic model to generate scale-free networks with the mechanism of preferential attachment. We denote the data set it generates as BA(N, m), where N is the size of the network and m is the number of initial ties that would be connected when a new node is added. Small World model is a random model with probability p to rewire and produce long range ties, it can be denoted as SW(N, K, p). CNNR model is modified from CNN [13] for generating social networks, especially online social networks. We denote it as CNNR(N, u, r), where $u(1 - r)$ is the probability to covert the potential edges into real ties. The averaged degree of the network it generates is approximately $2/(1 - u)$. The real-world data sets come from different fields. For example, CA-HepPh is a collaboration network from the e-print arXiv(http://www.arxiv.org) and covers scientific collaborations between authors of papers submitted to High Energy Physics [23]. NewOrleans is the Facebook network in New Orleans [24]. Email-Enron is an email communication network that covers all the email communication within a data set of around half million emails [25]. The basic properties

Table 1. Data Sets.

Data set	N	$|E|$
BA(20000, 10)	20000	199352
SW(20000, 10, 0.1)	20000	200000
CNNR(20000, 0.9, 0.04)	20000	187215
CA-HepPh	12006	118489
NewOrleans	63392	816886
Email-Enron	36692	183831

doi:10.1371/journal.pone.0136896.t001

of theses data sets we utilize in following experiments are listed in Table 1 and the real networks' download sources can be found in S1 Datasets.

As discussed before, establishing a new friendship may affect the entropy of the both ends. In the above networks, we characterize the relation between c_{ij} and $\Delta\epsilon(i, j)$ in the following steps: For each tie between i and j, we first obtain $\epsilon'(i)_j + \epsilon'(j)_i$ in the origin network; Secondly, we delete this tie and get $\epsilon(i)_j + \epsilon(j)_i$; Thirdly, the tie is restored. For different $\Delta\epsilon(i, j)$ for the same c_{ij}, we get the maximum, mean and minimum values, respectively. The change of entropy for other nodes in the network is not considered here for the reason that we assume the establishment of a tie between i and j is a personal activity with local information solely. As shown in Fig 2, in all networks, $\Delta\epsilon(i, j)$ decreases as c_{ij} grows, which is consistent with our above analysis, especially for the small world network in Fig 2b. At the start stage, the diverge between the maximum and mean of $\Delta\epsilon(i, j)$ is large, then it decays quickly as c_{ij} increases. It is also observed that for the nodes with tremendous common friends, building a new friendship between them may even lead to entropy loss. Note that except the small world network (Fig 2b), the deviation between the maximum and mean of $\Delta\epsilon(i, j)$ can be very large as c_{ij} is pretty small. It is because different from Poisson's distribution, the degree distribution of the real networks and BA model are power-law. And the existing of hub nodes with extremely large degrees in those networks might possess very high information entropy gain but low common neighbors (like a star), and therefore the variance of $\Delta\epsilon(i, j)$ can be very large as c_{ij} is tiny.

To sum up, the empirical results testify our statement further that increment of homophily would reduce the information entropy gain, which indicates a competition between the two evolving rules.

Positiveness

The growing of a social network could be simply regarded as establishing new ties among individuals. From the perspective of information entropy maximization, a tie should be established to gain more entropy for both ends. Therefore, we could distinguish the tie that makes the entropy of its ends gain as the positive tie, while the one that leads to entropy loss as the negative tie. Then we define the positiveness of the social network as the fraction of positive ties, which is denoted as τ. Larger τ means more ties in the network are established to increase their ends' entropy gain. As shown in Table 2, we list τ of the real-world network, where c is the clustering of the network. It is interesting that for the network with higher c, its τ is lower generally. We also investigate this finding on the network with various clusterings generated by BA and Small World models. For the BA model, we employ the method of tuning clustering while keeping its degree distribution stable [26, 27]. We only perform experiments of tuning the clustering on BA(1000,4), because it is too much time consuming for BA(20000,10). For the model of Small World, we just vary p. As shown in Fig 3, for both of models, the positiveness of
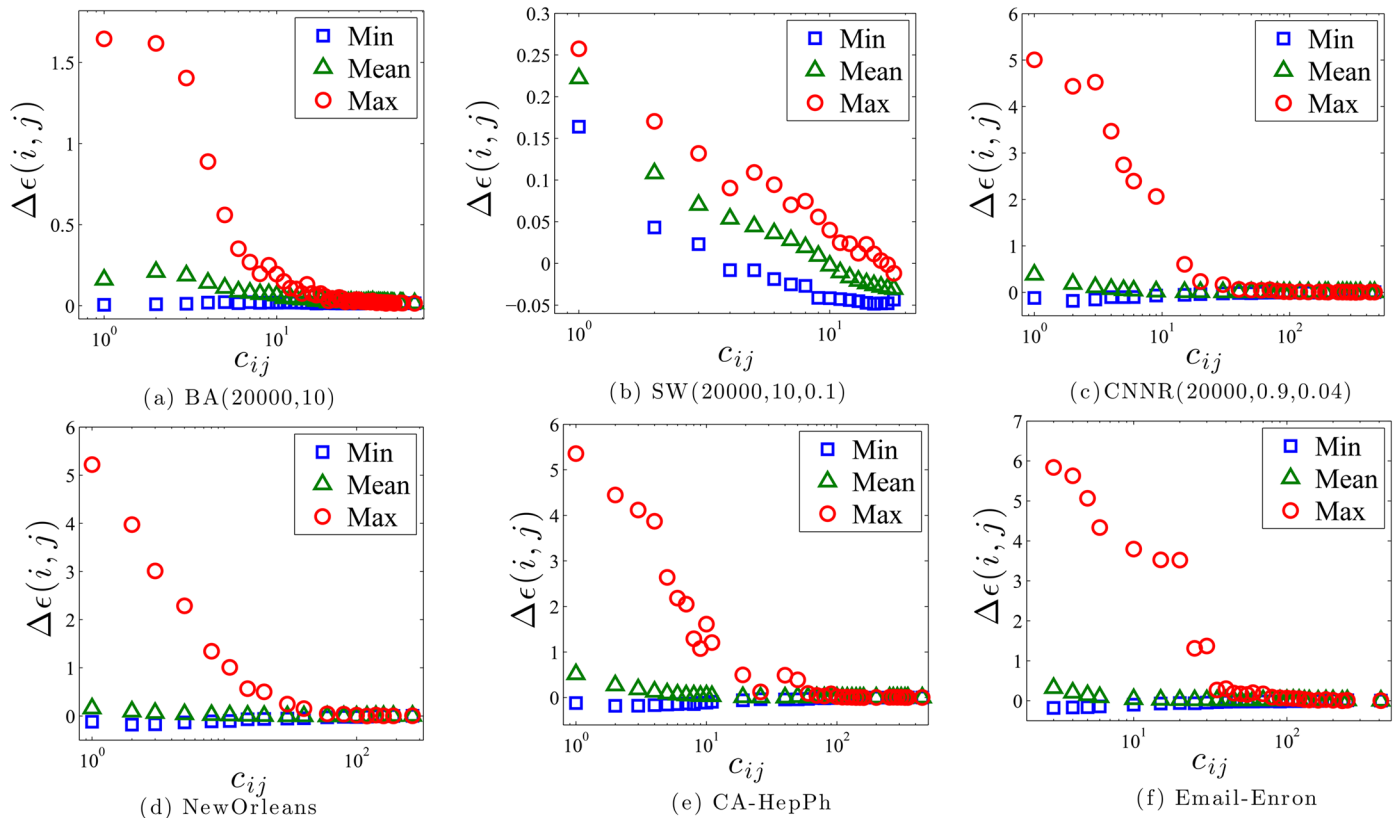
Fig 2. Empirical results from data sets. The results are consistent with the theory that increment of common friends would decrease the information entropy gain, especially for the maximum. Particularly, it should be also noted that as predicted by the analytical results, the averaged decay of $\Delta\epsilon(i,j)$ is very small in some cases, as shown in Fig 2d. Note that there are several outliers for the maximum $\Delta\epsilon(i,j)$, like in Fig 2c, which are produced by the noise in statistics. While the global trend of decrement with c_{ij} in all networks is still significant.

doi:10.1371/journal.pone.0136896.g002

Table 2. τ of the real-world networks.

Data set	τ	c
NewOrleans	0.70	0.22
Email-Enron	0.56	0.50
CA-HepPh	0.50	0.61

doi:10.1371/journal.pone.0136896.t002

network decreases as c grows. In fact, the clustering of the network could be rewritten [28] as

$$c = \frac{1}{|V|} \sum_{\forall(i,j)\in E} \frac{c_{ij}}{\binom{k_i}{2}}.$$

For this reason, with respect to the rule of homophily, a new tie added preferentially between nodes with overlapped friends would also lead to new triangles constructed in local structure. That is to say, the clustering of the network, i.e., c, would be increased when its evolution is driven by the homophily. Because of this, homophily dominated evolution leads to the decrement of τ. However, with respect to the information entropy maximization, the new tie is
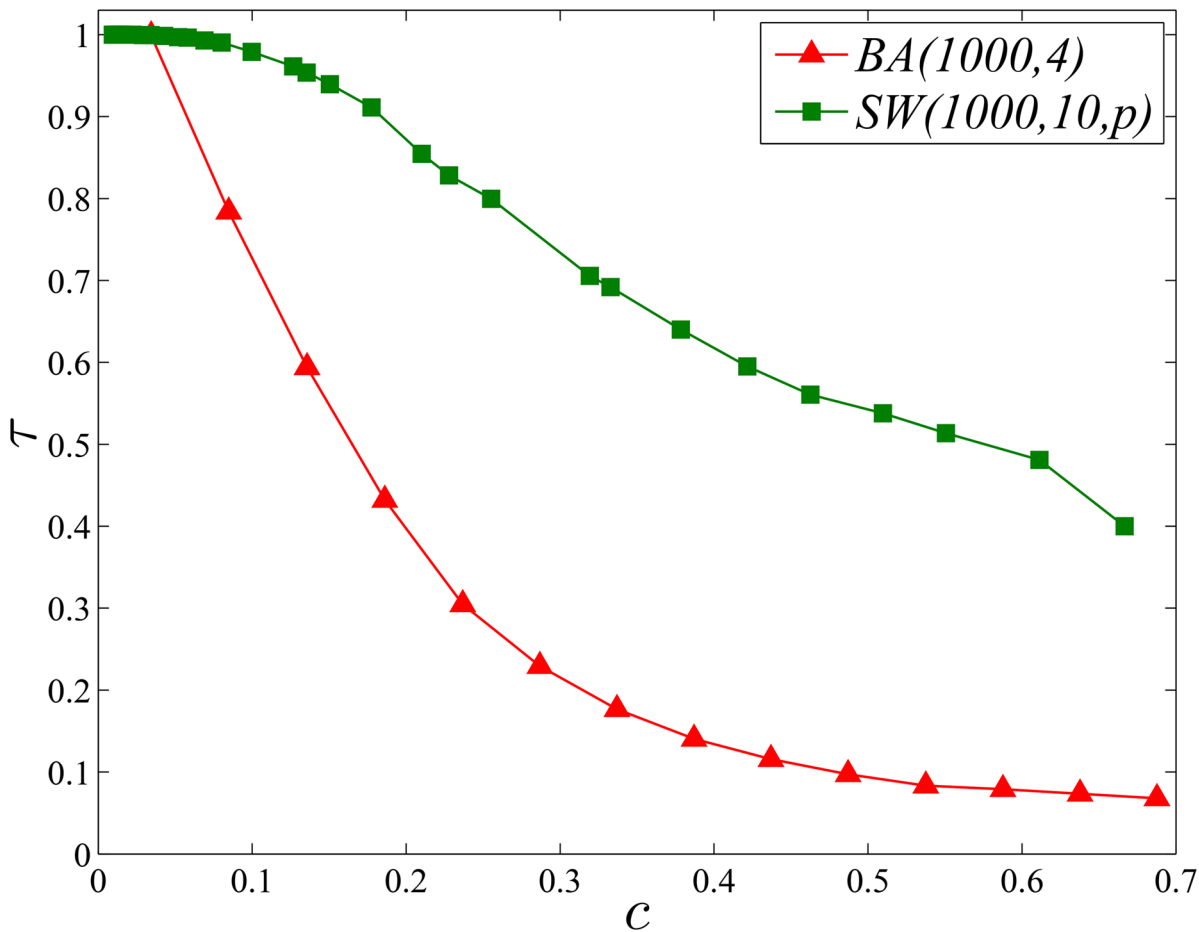
Fig 3. τ of the network varies as c increases.

established to increase the diversity of the information source and gain more entropy, which would improve τ by importing more positive ties.

The strength of a social tie can be defined as the number of overlapped friends between its ends. For example, the strength of a tie between i and j could be defined as $w_{ij} = c_{ij}/(k_i - 1 + k_j - 1 - c_{ij})$ [29–31], where lower w_{ij} stands for a weak tie. It is obvious that if i and j share a lot of common friends, the strength of the tie between them is strong. Conventionally, it is thought that the weak tie is helpful in getting the new information [32], while the strong tie means the relationship is trustful [19]. Therefore, based on the above discussion, it seems that the evolution supervised by homophily could lead to generations of strong ties in the network, because it renders the network highly clustered. In order to validate this, we observe the cumulative distribution function(CDF) of w_{ij} for each tie in the network. As shown in Fig 4, as c of the network decreases, the CDF curve moves to the left, which indicates the increment of the fraction of weak ties [33]. It validates our conjecture that in both synthetic and real-world data sets, highly clustered networks caused by homophily contain more strong ties, while the ones with lower clusterings contain more weak ties, which are produced by the law of maximum information entropy.
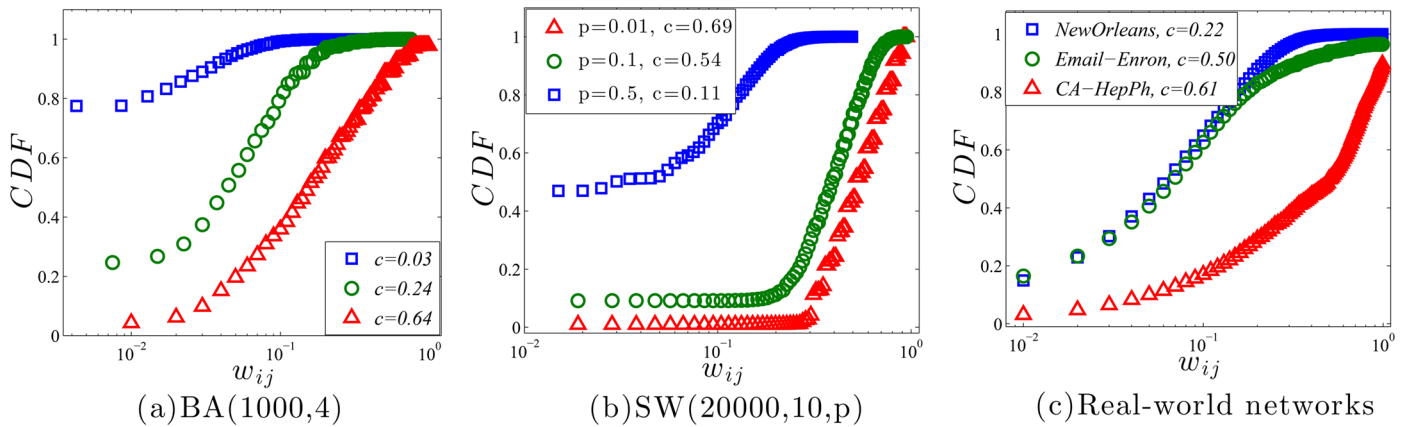
Fig 4. CDF of w_{ij} for various c.

Competition Model

A simple toy model is built to further demonstrate and understand the competition between information maximization and homophily in social networks' evolution. In this model, we simply assume the network starts to evolve from a sized-fixed but extremely sparse BA network and new links are added based on their scores, which can be calculated as

$$\lambda \Delta \epsilon(i,j) + (1 - \lambda) c_{ij}$$

at the initial stage, where i and j are a pair of non-connected nodes in the starting graph and $0 \leq \lambda \leq 1$ is a parameter to tune the role of information maximization in the generation of new ties. Intuitively, as λ getting close to 1, new links that can bring high information entropy gain (represented by $\Delta \epsilon(i,j)$) will be preferentially selected, while contrarily, as λ getting close to 0, links with high homophily (represented by c_{ij}) will be first established. Note that in order to make $\Delta \epsilon(i,j)$ and c_{ij} comparable, we normalize them by dividing their maximum values respectively.

As can be seen in Fig 5, the competition between information maximization and homophily can be well reproduced through our toy model. Specifically, we can find that when λ grows, the average clustering (denoted as c) of the network begins to increase until arriving at the maximum value, because small λ indicates that new links are mainly generated between nodes with high c_{ij} and many triangles might emerge locally. While regarding to τ, the fraction of positive ties, it first decreases until to its minimum and then begins to increase steadily, because as λ grows, the rule of information maximization will select more and more positive links from the candidate and the local clustering will be broken by weak ties of high entropy gain. It is also consistent with our previous finding that τ is negatively correlated with c.

Meanwhile, from Fig 5, we also notice that there exists a critical λ_c for τ and c, respectively. For instance, the average clustering of the network will arrive at the maximum as $\lambda_c = 0.35$, while the positiveness of the network arrives at the minimum when $\lambda_c = 0.6$. The first critical value suggests that as $\lambda < 0.35$, the rule of homophily dominates the evolution of the social network, while the second critical value indicates that as $\lambda > 0.6$, the rule of information maximization begins the dominate the formation of new ties in the evolution. However, as $0.35 \leq \lambda \leq 0.6$, both of the rules coexist and function simultaneously in the evolution.
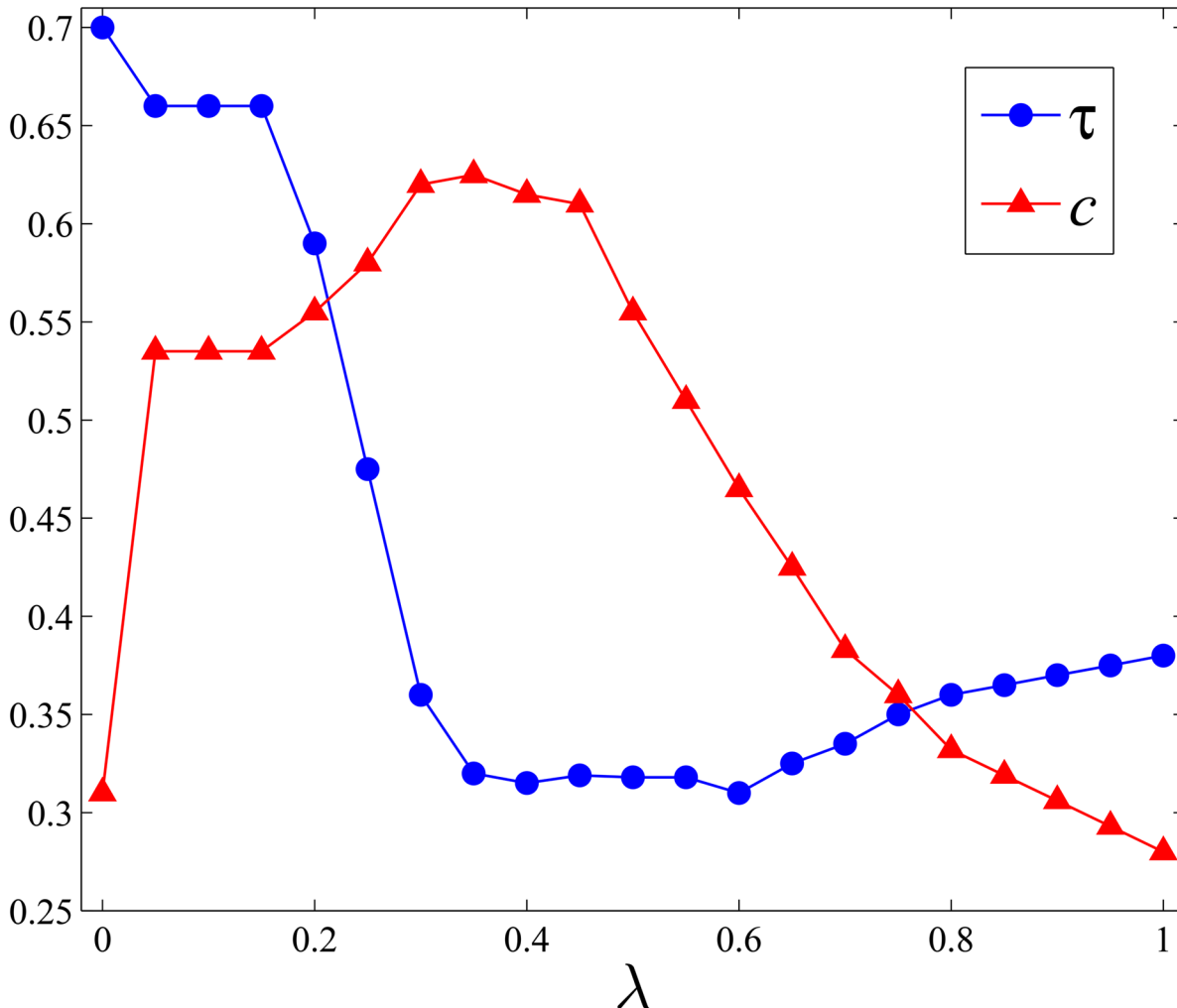
Fig 5. Evolutions of different λ. The size of the network is 1000 and the initial average degree is 4. 10000 new links have been added to guarantee the stability of the results for each λ.

doi:10.1371/journal.pone.0136896.g005

Moreover, as reported in Table 2, the average clustering and positiveness of real networks employed here are around 0.5, which means λ for real networks we used is smaller than 0.35 (as seen in Fig 5) and the homophily mainly drives the evolution and the rule of information entropy functions limitedly.

To sum up, the toy model developed here can well demonstrate the competition between the rules of information maximization and homophily and it also provides us a way to determine which rule plays the dominant role in the evolution by evaluating the value of λ from the views of clustering and positiveness. However, this model ignores the coming of new nodes and the ties' score is only determined by the initial status, which indeed needs further enhancement in the future work.

Conclusion and Future Work

In summary, both theoretical analysis and experimental results show that the rule of homophily is competing with the law of information entropy maximization in social networks.

Moreover, the rule of homophily driven evolution makes the network highly clustered and increases the certainty of the information source for a node. Contrarily, the rule of maximum entropy leads to the diversity of information sources. Based on the definition of weak ties, we can conclude that the rule of maximum information entropy leads to the generation of weak ties in the network, while the homophily produces strong ties between nodes with overlapped friends. Corresponding to the fact that both the weak and strong ties coexist in the network, we conjecture that both of the evolving rules might coexist in growth of the social networks. Therefore, in the view of maximum information entropy, the social network is not efficient, however, it owns many strong ties which may deliver trust information. We also develop a toy model to demonstrate the competition of different evolving rules and it can help to distinguish the different roles of different laws in real networks. Our findings could provide insights for modeling social network evolution as a competition of different rules.

This study has inevitable limitations. First, too many factors are neglected in the competition analysis and a more sophisticated and predicable framework is necessary. For example, given the tremendous development of the online social network, the cost of social activity in the epoch of the Internet continues to decrease [19, 20], and because of this, we neglect the cost of establishing ties of different strengths for simplifying the analytical framework. While in the real world, the strategic activity can be constrained by the personal cognition limit and social cost [34, 35] and the Dunbar's number [36] still exists in the online social network [20, 37, 38]. Hence in the future work, we would take the cost of establish different ties into consideration and build an evolution model of social networks based on the competition of strong and weak ties. Second, the empirical evidence from evolution of real networks is missing. So collecting fine-grained evolving trajectories of real social networks can be another interesting direction in our future work.

Supporting Information

S1 Equation. Expansion of $\Delta\epsilon(i)_j$.
(PDF)

S1 Datasets. The datasets download location. All the real-world data sets employed in this paper is publicly available and they can be downloaded freely from the following permanent location in figshare.com: http://dx.doi.org/10.6084/m9.figshare.1512836.
(PDF)

Author Contributions

Conceived and designed the experiments: JZ KX. Performed the experiments: JZ XL. Analyzed the data: JZ XL KX. Contributed reagents/materials/analysis tools: JZ XL. Wrote the paper: JZ XL KX.

References

1. Newman MEJ. The Structure and Function of Complex Networks. SIAM Rev. 2003; 45(167).

2. Zhou H, Lipowsky R. Dynamic pattern evolution on scale-free networks. Proc Natl Acad Sci. 2005; 102 (29):10052–10057. doi: 10.1073/pnas.0409296102 PMID: 16006533

3. Zhang W, Pan Y, Peng P, Li J, Li X, Li A. Global core and galaxy structure of networks. Sci China Inf Sci. 2014; 57:072101. doi: 10.1007/s11432-013-4930-6

4. Newman MEJ, Park J. Why social networks are different from other types of networks. Phys Rev E. 2003; 68(3):036122. doi: 10.1103/PhysRevE.68.036122

5. Jin EM, Girvan M, Newman MEJ. Structure of growing social networks. Phys Rev E. 2001; 64 (4):046132. doi: 10.1103/PhysRevE.64.046132

6. McPherson M, Smith-Lovin L, Cook JM. Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 2001; 27:415–444. doi: 10.1146/annurev.soc.27.1.415

7. Zhang W, Nie L, Jiang H, Chen Z, Liu J. Developer social networks in software engineering: construction, analysis, and applications. Sci China Inf Sci. 2014; 57:121101. doi: 10.1007/s11432-014-5221-6

8. Kibanov M, Atzmueller M, Scholz C, Stumme G. Temporal evolution of contacts and communities in networks of face-to-face human interactions. Sci China Inf Sci. 2014; 57:032103. doi: 10.1007/s11432-014-5067-y

9. Goldenberg J, Levy M. Distance Is Not Dead: Social Interaction and Geographical Distance in the Internet Era. arXiv:09063202v2. 2009;.

10. Hu Y, Wang Y, Li D, Havlin S, Di Z. Possible Origin of Efficient Navigation in Small Worlds. Phys Rev Lett. 2011; 106(10):108701. doi: 10.1103/PhysRevLett.106.108701 PMID: 21469842

11. Kossinets G, Watts DJ. Empirical analysis of an evolving social network. Science. 2006; 311(5757):88–90. doi: 10.1126/science.1116869 PMID: 16400149

12. Davidsen J, Ebel H, Bornholdt S. Emergence of a Small World from Local Interactions: Modeling Acquaintance Networks. Phys Rev Lett. 2002; 88(12):128701. doi: 10.1103/PhysRevLett.88.128701 PMID: 11909506

13. Vázquez A. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys Rev E. 2003; 67(5):056104. doi: 10.1103/PhysRevE.67.056104

14. Yuta K, Ono N, Fujiwara Y. A gap in the community-size distribution of a large-scale social networking site. arXiv:physics/0701168. 2007;.

15. Toivonen R, Onnela JP, Saramäki J, Hyvönen J, Kaski K. A model for social networks. Physica A. 2006; 371(2):851–860. doi: 10.1016/j.physa.2006.03.050

16. Holme P, Kim BJ. Growing scale-free networks with tunable clustering. Phys Rev E. 2002; 65 (2):026107. doi: 10.1103/PhysRevE.65.026107

17. Liben-Nowell D, Kleinberg J. The link prediction problem for social networks. In: the twelfth international conference on Information and knowledge management. CIKM'03; 2003. p. 556–559.

18. Lü L, Zhou T. Link prediction in complex networks: A survey. Physica A. 2011; 390:1150–1170. doi: 10.1016/j.physa.2010.11.027

19. Mislove A, Marcon M, Gummadi KP, Druschel P, Bhattacharjee B. Measurement and analysis of online social networks. In: the 7th ACM SIGCOMM conference on Internet measurement. IMC'07; 2007. p. 29–42.

20. Ahn YY, Han S, Kwak H, Moon S, Jeong H. Analysis of Topological Characteristics of Huge Online Social Networking Services. In: Proceedings of the 16th International Conference on World Wide Web. WWW'07. New York, NY, USA: ACM; 2007. p. 835–844.

21. Barabási AL, Albert R. Emergence of Scaling in Random Networks. Science. 1999; 286(5439):509–512. doi: 10.1126/science.286.5439.509

22. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998; 393:440–442. doi: 10.1038/30918 PMID: 9623998

23. Leskovec J, Kleinberg J, Faloutsos C. Graph Evolution: Densification and Shrinking Diameters. ACM Trans Knowl Discov Data. 2007; 1(1). doi: 10.1145/1217299.1217301

24. Viswanath B, Mislove A, Cha M, Gummadi KP. On the Evolution of User Interaction in Facebook. In: WOSN'09; 2009. p. 37–42.

25. Leskovec J, Kleinberg J, Faloutsos C. Graphs over time: densification laws, shrinking diameters and possible explanations. In: the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining. KDD'05; 2005. p. 177–187.

26. Kim BJ. Performance of networks of artificial neurons: The role of clustering. Phys Rev E. 2004; 69 (4):045101. doi: 10.1103/PhysRevE.69.045101

27. Ma X, Huang L, Lai YC, Zheng Z. Emergence of loop structure in scale-free networks and dynamical consequences. Phys Rev E. 2009; 79(5):056106. doi: 10.1103/PhysRevE.79.056106

28. Ribeiro B, Towsley D. Estimating and sampling graphs with multidimensional random walks. In: the 10th annual conference on Internet measurement. IMC'10; 2010. p. 390–403.

29. Onnela JP, Saramaki J, Hyvonen J, Szabo G, Lazer D, Kaski K, et al. Structure and tie strengths in mobile communication networks. Proc Natl Acad Sci. 2007; 104(18):7332–7336. doi: 10.1073/pnas.0610245104 PMID: 17456605

30. Cheng XQ, Ren FX, Shen HW, Zhang ZK, Zhou T. Bridgeness: a local index on edge significance in maintaining global connectivity. Journal of Statistical Mechanics: Theory and Experiment. 2010; 2010 (10):P10011. doi: 10.1088/1742-5468/2010/10/P10011

31. Zhao J, Wu J, Xu K. Weak ties: Subtle role of information diffusion in online social networks. Phys Rev E. 2010; 82(1):016105. doi: 10.1103/PhysRevE.82.016105

32. Granovetter MS. The Strength of Weak Ties. University of Chicago Press; 1974.

33. Zhao J, Wu J, Feng X, Xiong H, Xu K. Information Propagation in Online Social Networks: A Tie Strength Perspective. Knowledge And Information System(KAIS). 2012; 32:589–608. doi: 10.1007/s10115-011-0445-x

34. Pollet TV, Roberts S, Dunbar R. Use of social network sites and instant messaging does not lead to increased social network size, or to emotionally closer relationships with offline network members. Cyberpsychology, Behavior, And Social Networking. 2011; 14(4):253–258. doi: 10.1089/cyber.2010.0161

35. Bao P, Shen HW, Chen W, Cheng XQ. Cumulative Effect in Information Diffusion: Empirical Study on a Microblogging Network. PLoS ONE. 2013; 8(10):e76027. doi: 10.1371/journal.pone.0076027 PMID: 24098422

36. Dunbar R. Grooming, Gossip, and the Evolution of Language. Harvard University Press, Cambridge, MA; 1998.

37. Zhao J, Wu J, Guannan L, Tao D, Xu K, Chunyang L. Being rational or aggressive? A revisit to Dunbar's number in online social networks. Neurocomputing. 2014; 142:343–353. doi: 10.1016/j.neucom.2014.04.025

38. Golder SA, Wilkinson DM, Huberman BA. Rhythms of social interaction: messaging within a massive online network. Communities and Technologies. 2007;p. 41–66.