A simple phenomenological model for time evolution of social networks

J. Jiang$^1$, Q. A. Wang$^2$, W. Li$^3$ and X. Cai$^3$

$^1$Research Center of Nonlinear Science and College of Mathematics and Computer Science of Wuhan Textile University, Wuhan, 430200, China
$^2$Institut Supérieur des Matériaux et Mécaniques Avancées 44, Avenue F.A. Bartholdi, 72000 Le Mans, France
$^3$Complexity Science Center, Institute of Particle Physics, Hua-Zhong (Central China) Normal University, Wuhan 430079, China

E-mail: jiangj2007010209@gmail.com

Abstract. Inspired by the maxim "long union divides and long division unites", a phenomenological model with the simplification of real social networks is proposed to explore the evolutionary features of these networks composed of the entities whose behaviors are dominated by two events: union and division. The nodes are endowed with some attributes such as identity, ingredient, richness, age and internal diversity, which determine collectively the evolution in a probabilistic way. Through the local interaction of two events, a stationary state of network is reached as a constant amount of nodes survive with no more event happened in the network, like a situation of tripartite confrontation. Besides, the number of survived nodes and the speed of network evolution can be controlled by two parameters.

1. Introduction

This work is a network modeling of the social systems composed of a large number of entities in interaction whose existence is dominated by two major events: union and division. Union means the unification of two or more nodes (of the network) into one. Division is the inverse process: one node splits into several ones. Union and division are also the most visible social, economic, cultural and political phenomena in the course of the development of many composite systems. Different countries, political or economic groups can be unified into one country or group like the reunion of East and West Germany in 1990. There are also plenty of examples of division of these unities like the disintegration of Soviet Union in 1991. All of them are the consequence of the interplay of a pair of opposite tendencies in the evolutionary systems, and the recurrent character of these phenomena is well summarized in one Chinese maxim "long union divides and long division unites".

Union and division are veritable complex processes in which many factors are responsible for the consequences. A good example of this complexity is the evolution of different economic groups in the same field through the interaction including communication, influence, competition or cooperation etc. The evolution is composed of unification and splitting of groups and is under very complicated and uncertain influence of economic, political, scientific and technological systems as well as of many accidental elements such as natural disasters, wars, environmental changes and so on. It is for this reason that the most suitable description of the stochastic
evolution of this kind of systems is the probabilistic modeling, taking into account as many of
the involved factors and interaction mechanisms as possible.

Here, we present a phenomenological modeling of networks in which union and division of
nodes are two main events determined in a probabilistic way by the nature and some general
features of the nodes. One of the aims is to see what would be the destiny and the evolution
characters of a network composed of nodes endowed with some general attributes allowing
and influencing the union and division under some given conditions, without entering into the
fundamental principles, interactions and true mechanisms of the dynamics of the systems. The
definition of a network in the present work is inspired by the previous works on cultural networks
[1, 2, 3, 4, 5, 6, 7, 8, 9, 10], where a culture is represented by a vector whose components are
quantified by a limited number of values characterizing the cultural features. The vectorial
representation of the nodes has been used in many other systems. For examples, in an economic
network formed by companies (nodes), each company is labelled with several essential features
like the richness or capital, the scale, the business scope, the survival time or brand image, and
so on. In a political network composed of parties or groups, each group has its own political
standpoint. In addition, the phenomenological model of union and division has been applied
recently to a dual modeling of political opinion networks [11], in which union and division are
two events happening in the subnetwork of political parties.

In the present model, we always have in mind some questions about, for instance: whether
or not it is possible for all nodes of a given network to merge into one node under reasonable
conditions. Or what will be the equilibrium or stationary state of network in which the number
of nodes is relatively unchangeable, like the mode of tripartite confrontation? And what are the
necessary conditions, if any, for these evolutionary behaviors? The reasons for considering these
questions are that the globalization of economy, culture and politics is becoming an increasingly
hot topic and raises various opinions in the world [12, 13, 14]. In our previous work [15], under
some given conditions and definitions, we got some unexpected answers to these questions that
the evolution of the system either terminated in a final union with the disappearance of the
whole network and only a single node survived, or never ceased a nonequilibrium and periodic
evolution with fluctuating oscillation of network size and some other quantities. Here, in order
to keep finding the solution of the problems unsolved in our previous work, we extend our
work with more realistic considerations. The main modifications include the redefinition of the
quantity richness and the reintroduction of new rules of the merging and splitting processes,
which results in some new interesting findings.

Again, we would like to emphasize that in this probabilistic model, the mechanism of union
and division is not imposed deterministically on the evolution. At each step of evolution, union
or division may or may not happen, depending on the probability of occurrence (union, division
or nothing) calculated according to the nature and features of the nodes.

2. The model
Each node in the model represents a social entity such as a country, a culture, a political party,
or a company. A link between two nodes means the interaction between them. So in one given
network, each node denotes an individual entity when the object of entity is chosen, like in the
company network, different nodes denote different companies with same business scope; in the
culture network, nodes denote various cultures in the world. As stated above, some inherent
properties could be used to characterize the entity. One can label a country or a culture by
using some features like the language, religion, art, customs, economy, etc. A quantity named
"identity" is introduced for this purpose and each node is endowed with an identity that has
a position in the \( w \) dimensional space: \( \vec{c} = (c_1, c_2, c_3, \ldots c_w)(p \leq c_i \leq q) \) with \( i = 1, 2, \ldots w \)
where \( w \) is the number of features and we consider \( w = 3 \) in the simulation for simplification.
\( c_i \) is a random integer uniformly distributed, here we take \( p = 1, q = 10 \). The identity distance
between two nodes $i$ and $j$ is defined as:

$$d_{ij} = \sqrt{\sum_{k=1}^{3} \Delta c_k^2},$$

where $\Delta c_k = c_{ik} - c_{jk}$. This distance represents the difference between entities and similar entities have small distance between them.

As different entities in the system considered have in general different levels and power of development, so the quantity "richness" is introduced to be an important attribute of an entity. For a culture, this may stand for the size and the power of its content (religion, art, language, literature, economy, education, science and technology, etc.), the population of its carriers and so on. For a country, it could indicate economic, political, cultural, military power, living conditions, as well as the population, etc. A larger value of richness means a higher level of development and power. Compared to the time evolution of richness defined as the product of two attributes, the ingredient and age in Ref.[15], here, we reconsider and generalize it as follows:

$$\vec{r}_i(t) = \alpha_1 r_i(t) \vec{1} + \alpha_2 r_i(t) \vec{2} + \cdots + \alpha_w r_i(t) \vec{w},$$

where $\alpha_i (i = 1, 2, \ldots, w)$ is the weight of $i$-th feature and $\sum_{i=1}^{w} \alpha_i = 1$. This form is inspired by the concept of the rankings in society. For example, for an university, the University Rankings Index is one potential way to directly reflect the development or richness of one university. Like the famous Times University Ranking, five factors ($w = 5$) are contained as teaching, research, citations, industry income, and the international outlook with weight of 0.3, 0.3, 0.3, 0.025 and 0.075, respectively, and Harvard University had the highest richness value among 200 universities with top index 96.1 in the ranking 2010 [16]; for a company, the financial power (or turnover, wealth) could be labelled as richness for simplification with $w = 1$ such as Fortune 500 Index and the company "Wal-Mart Stores" got the highest richness value with the turnover 421849 million dollars among the Global Top 500 Companies in 2010 [17]; for a political party, the public support rate might be a suitable candidate for its richness. So various rankings of different entities like university, company, city competitiveness, etc, to some extent, could play the role of function of the richness in this model. Here, we take into account the age ($a$), the degree ($k$), and the ingredient ($I$) in the definition of richness for simplification as follows:

$$\vec{r}_i(t) = \alpha_1 a_i(t) \vec{1} + \alpha_2 k_i(t) \vec{2} + \alpha_3 I_i(t) \vec{3},$$

where $\alpha_i (i = 1, 2, 3)$ is weight value, and the age of a node is the time period from its birth to the present time, and the degree $k$ or number of connections describes the communication state, the openness, the ability and will be of giving or receiving information of the node, and the ingredient $I$ of a node is defined as the number of nodes that have been merged.

Next, we introduce two main dynamic processes in the network: merging and splitting. For two nodes randomly chosen, $i$ and $j$, $j$ is one of the nearest neighbors of $i$, they can merge into one new node with probability $p_m$ proportional to the similarity or overlap (number of common identity features $c_{li}$, $l = 1, \cdots, w$) between the identity of two nodes. The probability is defined as:

$$p_m(i, j) = \frac{1}{w} \sum_{l=1}^{w} \delta(c_{li}, c_{lj}),$$

where the $\delta$ function takes value 0 or 1 when $c_{li}$ is equal or not equal to $c_{lj}$. This expression pays more attention to the similarity between the identity of two nodes since on the intuitive level, the more similar two nodes are, the more likely they are close to each other and merge
together, as being said: *Birds of a feather flock together.* For example, two kinds of western or eastern cultures have more likely to merge than one eastern culture and one western culture do. If two nodes have totally different or identical features of identity, the probability of their merging will be zero or one. Otherwise, it will be \(1/w \times p_m < 1\). After the merging of two nodes, the new node \(n\) will be created with age \(a = 1\) as a composite one characterized by the ingredient \(I\). Obviously, the maximum ingredient is equal to the initial size of network \(N\) and will be conserved during the dynamic evolution. The richness of the new node evolves as Eq.(3) and its identity takes the average value of that of two merged nodes as following:

\[
\vec{c} = \frac{c_{1i} + c_{1j}}{2} + \ldots + \frac{c_{wi} + c_{wj}}{2}.
\]

The degree of the new node \(n\) is given by

\[
k_n = k_i - 1 + k_j - 1 - N_{\text{common}},
\]

where \(N_{\text{common}}\) is the number of common neighbors of node \(i\) and \(j\). In other words, the topological evolution of the network after node merging behaved as all neighbors of merged nodes \(i\) and \(j\) will be the neighbors of the new composite node \(n\).

For the process of splitting in the model, a randomly selected composite node with identity \(I (I \geq 2)\) could split in two ways. One is full or complete splitting generating \(I\) new nodes, has been studied in our previous work. The other proposed in the present work is partial splitting generating \(s\) new nodes \((2 \leq s \leq I - 1)\), in order to complete the mechanism of splitting and compare the effect on network dynamics caused by splitting methods. The splitting probability \(p_s\) is given from the new perspective of internal diversity as follows:

\[
p_s = \frac{1}{Z_s} \frac{d_I}{I(I-1)} = \frac{1}{Z_s} \frac{2 \sum_{i,j\in I} d_{ij}}{I(I-1)},
\]

where \(Z_s\) is the normalization constant, normalized over all composite nodes in the network, \(d_I\) is the average identity distance among \(I\) nodes merged by the composite node, which is also named as internal diversity, and \(i, j\) are the indexes of \(I\) nodes. The larger internal diversity \(d_I\) of one composite node is, the more likely it will fragment. In other words, the more internal solidarity one group has, the more difficult it disintegrates. After splitting, the identity of the new created nodes has the same dimensional number and the same range as the split "mother" node, with randomly given values uniformly distributed. The age of new nodes is equal to one and the ingredient of them is \(I = 1\) for full splitting or a random integer between 1 and \(I - s\) for partial splitting with the conservation that the sum of \(s\) new nodes’ ingredients should be equal to \(I\). In addition, the richness of each new node also agrees with Eq.(3). The structure of the network after node splitting is that these new nodes are randomly connected with each other and with the neighbors of the split node with a constant probability \(p_{rc} = 0.9\).

It is worth noticing two conditions of time evolution in our simulation: 1) at a given time step, a composite node which has just been formed or a new node which has just split from a composite node must not be considered as the candidate for another merging or splitting; 2) during the random selection of the nodes, if the chosen node has \(I = 1\), the one and only possible event is merging. But if it is a composite one, merging and splitting are both possible. The decision will take place with the changeable "event probability" \(mors\) for each possible event in the present model, which is more realistic than the fixed probability 0.5 considered in the previous work.

The initial condition of the simulation is a number \(N = 150\) of nodes randomly connected with a constant probability \(p_c = 0.25\). The nodes remaining unconnected will be selected randomly at each time step and connected to other nodes according to the principle of preferential attachment.
As the weight of features defined in richness in Eq. (3) is unknown, the same weight value \( \alpha_i = 1/3 \) \((i = 1, 2, 3)\) is implemented in simulation. When \( t = 1 \), the age of each node is \( a = 1 \), and the ingredient \( I = 1 \).

3. Numerical results
Totally different from a quasi-periodic behavior and an alternative replacement of union and division state found in the previous result, an equilibrium state is shown in the present model where the final size of network is stationary with time evolution. After the competition of union and division, the final size of the network is similar with the mode of tripartite confrontation and no more dynamic evolution happens. Details of the time evolution of some quantities are described below.

3.1. Cumulative network size evolution
The time evolution of cumulative network size with two splitting methods and different event probability \( mors \) is shown in Fig. 1. The cumulative network size at time step \( t \) is defined as \( \sum_{i=1}^{t} n(i) \), where \( n(i) \) is the size of network (the number of nodes in network) at time step \( i \). Since the final time step is various for network size evolution at stationary state with different event probability \( mors \), the X-axis time scale of the figure is rescaled with the same range from 0 to 1 by dividing the final time step. In the left side of Fig.1, with partial splitting process, the cumulative network size decreases with the time growing for each event probability \( mors \) with small variation. The merging and splitting processes are responsible for the oscillatory behavior of the cumulative size evolution. With the increase of event probability \( mors \) from 0.1 to 0.9, or the decision to split for a composite node assumed in the simulation, the final network size increases from 15 to 36 approximately and the average final network size over different event probability \( mors \) is around 26. In addition, the time step needed to reach the stationary state is sharply shortened with the increase of \( mors \) during the evolution. In the right side of Fig.1, with full splitting process, the cumulative network size has similar tendency in time, but with large variation. The fluctuating range of final network size becomes much narrower between 24 and 34 approximately with event probability varying from 0.1 to 0.9 and the average final network size over different event probability \( mors \) is around 27.3 that is larger than that with partial splitting. Furthermore, the whole network needs more evolution time to reach the stationary state as more new nodes are split from the composite node with full splitting method compared to the case with partial splitting.

From the comparison between two splitting methods, we can draw a conclusion that the network can evolve to stationary state more quickly with partial splitting method, where larger stable network size could be found if full splitting method is used.

3.2. Richness distribution
We studied the richness distribution (represented by empty circle) and made a comparison with empirical data in Fig. 2 in which the richness is normalized by dividing the maximum value of each data. The data sets include (1) the Global Urban Competitiveness Rankings Index 2007/2008 [18] (data 1, represented by red circle): for each city competitiveness analysis over 500 main cities in the world, it is presented for seven sectors of the economy, such as industrial structure, human resources, the living environment, and so on. Each of these sectorial indices is the result of data for four to seven variables relating to aspects of each sector; (2) FT (full time) Weekly Median Pay (dollar per week) in 378 cities of USA in 2009 [19] (data 2, represented by black square); (3) the World City Knowledge Competitiveness Index in 2008 [20] (data 3, represented by blue square): it compares 145 regions across 19 knowledge economy benchmarks and five factors are selected for the analysis including human capital, knowledge capital, regional
Figure 1: The evolution of cumulative size of network with different splitting processes and event probability $mors$. In the left side of figure, the cumulative size of network with partial splitting method decreases with the time evolution up to one relatively stationary value with tittle variation. With the increasing of $mors$ from 0.1 to 0.9, the final stable size of network increases from 15 to 36 approximately and less time step is needed to reach the stationary state. In the right side of figure with full splitting, similar behavior of cumulative size evolution is found over the time. The main differences from the left side are: 1) the fluctuating range of final stable size of network becomes narrow between 24 and 34 with event probability $mors$ varying from 0.1 to 0.9; 2) more time steps are needed for network to reach the stationary state with the increase of event probability.

3.3. Distribution of the split nodes number
From the introduction of node splitting process in Section 2, we know that for the full splitting process, the number of split nodes from the composite node is equal to the ingredient $I$ of composite node, and for the partial splitting process, the number of split nodes is between 2 and $I - 1$. We record the number of split nodes during the whole evolution with given event probability $mors = 0.5$ in the simulation and show its probability distribution with two node splitting methods in Fig. 3. We find that 1) two distributions are power laws with fat tails in log-log plot, which suggests that during the time evolution, most of composite nodes with small ingredients split, and there are only a few of them with large ingredient, since these nodes have large probability to split. In the meanwhile, it also indicates that the probability distribution of ingredient matches the power law distribution; 2) power laws have two scaling exponents -1.4 and -1.6, fitted by two straight lines, corresponding to full and partial splitting method, respectively. Since for one composite node, partial splitting always produces less new nodes than full splitting does, so the scaling exponent for partial splitting is smaller than that for full splitting, and the maximum number of split node for former is also smaller than that for latter.
Figure 2: Comparison of the richness distribution between the model simulation and the real data including Top 500 Global Urban Competitiveness Rankings 2007/2008 (red circle, data 1)[18], the FT (full time) Weekly Median Pay in 378 cities of USA in 2009 (black square, data 2)[19], and the World Knowledge Competitiveness Rankings in 2008 of 145 cities in the world (blue square, data 3)[20]. We find that the model and real data have the similar tendency that each distribution has a sharp peak at the point smaller than each richness mean value, and most of entities are in the low level of development: only a few of them develops well and have much power.

Figure 3: Distributions of the number of nodes split from composite node with full (empty circle) and partial (full circle) splitting methods. In log-log plot, distributions have the form of power law with fat-tail, with exponents -1.4 (full splitting) and -1.6 (partial splitting) fitted by two straight lines, respectively.

4. Conclusions and discussions
In the present paper, we have extended the phenomenological model of network evolution in our previous work with some important modifications including the redefinition of richness and the evolution rule of merging and splitting. We generalized the definition of richness and found an analogy between it and the ranking index in real system through the comparison between the simulation of richness distribution and some real data. In addition, inspired by the saying
that birds of a feather flock together and the view of the internal diversity, we proposed new mechanisms of merging and splitting evolution functions.

As a result, some results totally different from previous work were found: for example, a stationary or equilibrium state of network size evolution has been obtained, in which a small amount of nodes survive in the network and no more merging or splitting process happens. The number of survived nodes depends on the value of event probability. Besides, the speed of network evolution is dependent on the node splitting method, full or partial splitting. At last, we found that the distribution of the number of node split from composite nodes follows the power law form with two splitting methods.

We want to stress that this model is still in its infancy and its applicability is limited with the social network where only the local interaction is dominated in the evolution, and more empirical data or accurate theoretical analysis is needed to support the simulation results. Thus this model has large flexibility to improve in the future work. For instance, global interaction or noise like random influence may take place in the process of merging and splitting and the features of some other quantities like the ingredient, the age and degree could be explored to enrich the findings of the model. We hope that this model, in spite of the simplification of the real world, could be a starting point for further work with more realistic and applicable considerations.

Acknowledgments
This work was supported by the National Science Foundation of China under grant No. 11405118.

References
[1] Axelrod R 1997 The dissemination of culture: A model with local convergence and global polarization, J. Conflict Res 41 203-26.
[2] Stanoev A, Smilkov D and Kocarev L 2011 Identifying communities by influence dynamics in social networks Preprint arXiv:1104.5247v1.
[3] Palla G, Dereny I, Farkas I and Vicsek T 2005 Uncovering the overlapping community structure of complex networks in nature and society Nature 435 814.
[4] Rosvall M and Bergstrom C T 2008 Maps of random walks on complex networks reveal community structure Proc. Natl Acad. Sci. USA 105 1118.
[5] Clauset A, Newman M E J and Moore C 2004 Finding community strucutre in very large networks Phys. Rev. E 70 066111.
[6] Lancichinetti A and Fortunato S 2011 Limits of modularity maximization in community detection Preprint arXiv:1107.1155v1.
[7] Newman M E J and Girvan M 2004 Finding and evaluating community structure in networks Phys. Rev. E 69 026113.
[8] Newman M E J 2006 Modularity and community structure in networks Proc. Natl. Acad. Sci. USA 103 8577.
[9] Kim B J, Trusina A, Minnlagen P and Sneppen K 2005 Self organized scale-free networks from merging and regeneration Eur. Phys. J. B 43 369.
[10] Alava M J and Dorogovtsev S N 2005 Complex networks created by aggregation Phys. Rev. E 71 036107.
[11] Wang R and Wang Q A 2011 Dual modeling of political opinion networks Phys. Rev. E 84 036108.
[12] Cornwell G H and Stoddard E W 2001 Global Multiculturalism: Comparative Perspectives on Ethnicity, Race and Nation (Rowman & Littlefield Publishers).
[13] Falk R 1999 Predatory Globalization: A Critique (Princeton: Princeton University Press)
[14] Chang Y F and Cai X 2007 Macro-Behaviour of agents’ opinion under influence of an external field Chin. Phys. Lett. 24 2430.
[15] Jiang J, Wang R, Michel P and Wang Q A 2013 Long division unites or long union divides: A model for social network evolution Chin. Phys. Lett. 30 038901.
[16] http://www.timeshighereducation.co.uk/world-university-rankings/
[17] http://www.fortunechina.com/fortune500/c/2011-07/07/content_62335.htm
[18] Ni P F and Kresl P K 2010 global urban competitiveness report 69-80.
[19] http://www.cforic.org/downloads.php
[20] Huggins R, Izushi H, Davies W and Shougui L 2008 World knowledge competitiveness index 7-13.