Chaos, Solitons and Fractals 110 (2018) 20–27

Contents lists available at ScienceDirect

Chaos, Solitons and Fractals

Nonlinear Science, and Nonequilibrium and Complex Phenomena

journal homepage: www.elsevier.com/locate/chaos

Frontiers

Evolving the attribute flow for dynamical clustering in signed networks

Hui-Jia Li [a,e,*], Zhan Bu [b], Yulong Li [a], Zhongyuan Zhang [c], Yanchang Chu [d,e], Guijun Li [a], Jie Cao [b]

[a] School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China
[b] Jiangsu Provincial Key Laboratory of E-Business, Nanjing University of Finance and Economics, Nanjing 210003, China
[c] School of Statistics and Mathematics, Central University of Finance and Economics, Beijing 100080, China
[d] Economics and Management College, Civil Aviation University Of China, Tianjin 300300, China
[e] The Research Center of Beijing-Tianjin-Hebei Civil Aviation Coordinate Development, Civil Aviation University Of China, Tianjin 300300, China

ARTICLE INFO

Article history:
Received 28 October 2017
Revised 7 February 2018
Accepted 8 February 2018

Keywords:
Signed networks
Attribute flow
Clustering algorithm
Dynamical systems
Convergence and divergence

ABSTRACT

In real networks, clustering is of great value to the analysis, design, and optimization of numerous complex systems in natural science and engineering, e.g. power supply systems ,modern transportation networks, and real-world networks. However, the majority of them simply pay attention to the density of edges rather than the signs of edges as the attributes to cluster, which usually suffer a high-level computational complexity. In this paper, a new rule is proposed to update the attributes flow, which can guarantee network clustering reach a state of optimal convergence. The positive and negative update rule we introduced, represent the *cooperative* and *hostile* relationship, and the attribute configuration will convergence and one can identify the reasonable cluster configuration automatically. An algorithm with high efficiency is proposed: a nearly linear relationship is found between the time complexity and the size in sparse networks. Finally, we conduct the verification of the algorithmic performance by a representative simulations on Correlates of War data.

1. Introduction

In the real world, social networks are used in modeling of numerous complex systems [1–3]. We can use $G = \{V, E\}$ as a graph to define a social network, where the set of vertices is defined as $V = \{1, \ldots, n\}$, and the set of edges which connecting pairs of vertices is defined as E. An example is that vertices indicate agents/individuals ,and edges indicate relations/links between nodes in an interpersonal network. The signed social networks, which is used to describe the social networks with positive and negative links, where "friendship"as an example of "positive relationship" is denoted by positive links ,and equally a "negative relationship" such as "hostility" may be indicated by negative links [4–12]. An example can be cited in the Gahuku-Gama subtribes network is that the "political alliance relation" are represented by the positive links, while "political opposition relation" are represented by the negative links [13]. We can create the signed network by the relationships,and obtain more learning about the attributes of the social networks such as the cluster configuration [14,15] from the analysis and mining on it. Clusters in networks refer to the phenomenon when nodes of the network can be naturally grouped into sets such that each set is densely connected internally, thus dividing the network into smaller groups with dense internal connections and sparser external connections. The links whose density and signs combine to define the signed network clusters. Qualitatively, clusters in signed network are defined as subgraphs that the positive links are within the nodes in each group and the negative link are connected between the different groups of nodes [16,17]. For the discovery of the hidden cluster configuration, it is anything but simple to find the optimal and steady partition of the network.

Despite many clustering techniques have been presented for analysis in the field of complex network, the majority of them simply pay attention to the density of edges rather than the signs of edges as the attributes to cluster [18–20]. If an accuracy within acceptable limits is obtained based on comparison between the internal and external cohesion of a subgraph by the traditional heuristic methods, the complexity of computation is usually high-level [21,22]. In this paper, a new rule is proposed to update attributes, which can guarantee network clustering reach a state of optimal

* Corresponding author at: School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100080, China.

E-mail addresses: hjli@amss.ac.cn, lihuu2000@126.com (H.-J. Li), buzhan@nuaa.edu.cn (Z. Bu), caojie690929@163.com (J. Cao).

https://doi.org/10.1016/j.chaos.2018.02.009
0960-0779/

convergence. By introducing the positive and negative update rule to represent the *cooperative* and *hostile* relationship [23], one can prove the convergence and divergence of the attribute evolution in mean and find its conditions. Unlike the heuristic method, the number clusters is not specified by user, meanwhile, the reasonable partition can be automatically distinguished by it. An algorithm with high efficiency is proposed: a nearly linear relationship is found between the time complexity and the size in sparse networks. Finally, we conduct the verification of the algorithmic performance by a representative simulations on Correlates of War data.

2. Materials and methods

2.1. Signed network

An undirected connected social network without self loops can be considered as $G=(V,E)$, where the set of vertices is defined as $V=\{1,\dots,n\}$, and the set of edges which connecting pairs of vertices is defined as E. For a *signed network*, we mark these positive and negative edges with a plus sign "+" and a minus sign "−" respectively in E, where "+" indicates a collaboration or friend relationship and "−" indicates an opponent or enemy relationship. In order to analyse, E is divided into the positive and negative edge collections,which are separately denoted as E_{pst} and E_{neg}. The positive and the negative networks are separately denoted as $G_{pst}=(V,E_{pst})$ and $G_{neg}=(V,E_{neg})$, where $E_{pst}\cap E_{neg}=\emptyset$ and $E_{pst}\cup E_{neg}=E$. In General, it can be considered that the negative edge collection E_{neg} is nonempty. For signed network G, the following definition is used to define a k-clusters configuration.

Definition 1. Let's call $V=V_1\cup V_2\ldots\cup V_k$ is k-clusters configuration in a signed network G. The label of each edge within V_i or between different V_i is positive or negative, respectively. Here, every V_i is nonempty and any two different V_i satisfies disjoint.

Intuitively, cutting all the negative links, which makes distinguishing clusters in partitionable or balanced signed networks completed easily. The positive links will be merely included in the subgraphs we achieved and clusters will take shape. However, the discrimination assignment becomes extraordinary owing to some scenarios : 1) the signed social networks cannot be partitioned and 2) despite it is feasible to cut the signed social networks by partitioning, merely cutting negative links cannot achieve the optimal partition or the most natural partition. Actually before cutting out all negative links, some large subgraphs with some isolated nodes arising from large subgraphs in a great number. In order to identify more natural clusters, the steady partitions of subgraphs in an increasing number are supposed to be reasonably ignored as well as keeping some positive and negative links (or reasonably reduce their effect).

What we find especially interesting is that the definition of cluster configuration can be related to the balance theory in signed network [23,24].

Definition 2. A signed graph is defined as $G=(V,E)$. Then

(i) G is a *weak equilibrium* if it satisfies that there is an integer $k\geq 2$ and a k-way partition $V=V_1\cup V_2\ldots\cup V_k$, where $V_1,\dots,V_k$ are nonempty and disjoint with each other, in this way any edge between different V_i or within each V_i is negative or positive, respectively.

(ii) G is a *strong equilibrium* when it satisfies that itself is a weak equilibrium and $k=2$.

2.2. Attribute dynamics

The *attribute*, or *opinion* of the nodes can be denoted as a real and scalar value when the nodes initiate interactions. For each node at time k, its attribute vector is defined as $x(k)\in\mathbb{R}_n$. The attribute of nodes will be updated according to the situation that nodes interact with their collaborators or opponents. Particularly, only two nodes $\{i,j\}$ are chosen and the following rules are used to update their attribute at each time k.

• (*Positive Update Rule*) If $\{i,j\}\in E_{pst}$, we update the attribute of node $m\in\{i,j\}$ as

$$\begin{aligned}x_m(k+1)&=x_m(k)+\alpha(x_{-m}(k)-x_m(k))\\&=(1-\alpha)x_m(k)+\alpha x_{-m}(k),\end{aligned}\tag{1}$$

where $-m\in\{i,j\}\backslash\{m\}$ and $0\leq\alpha\leq 1$.

• (*Negative Update Rule*) If $\{i,j\}\in E_{neg}$, we update the attribute of node $m\in\{i,j\}$ as

$$\begin{aligned}x_m(k+1)&=x_m(k)-\beta(x_{-m}(k)-x_m(k))\\&=(1+\beta)x_m(k)-\beta x_{-m}(k),\end{aligned}\tag{2}$$

where $\beta\geq 0$

For the positive update rule, nodes update their attributes based on the previous attributes of nodes and of their neighbors which are regarded as a convex combination. The *cooperative* or *unsuspecting* relationships are shown naturally in this update. The positive update rule, considered as the attraction of the attributes, which tends to drive node attributes closer to each other.

On the other hand, there are controversy over the dynamics on the negative edges in the literature. Substantial efforts have been taken to characterize these *suspecting* or *hostile* relationships. The proposed negative update rule, is the *contrary* of the positive update rule, which enforces attribute differences between interacting nodes. Note that the negative update rule satisfies the following elaborations:

• Node i tries to *trick* her negative neighbors j, by turning to the opposite sign of her true attribute (i.e., $x_i(k)$ to $-x_i(k)$) before showing it to j;

• Node i distinguishes j as her negative neighbor and upon observing $x_j(k)$ which is j's true attribute, she attempts to get closer to the opposite view of j since $x_i(k+1)$ is a convex combination of $x_i(k)$ and $-x_j(k)$.

2.3. The mean convergence and divergence

Let define the (random)vector of attributes at time k resulting from the node interactions as $x(k)=(x_1(k),\dots,x_n(k)), k=0,1,\dots$. The initial attributes $x(0)$ is also denoted as x^0 and deemed to be featured in determinacy. In this section, we make a thorough investigation into the mean evolution of the attributes. We present the following definition.

Definition 3. (i) The *expected* attribute convergence is obtained if $\lim_{k\to\infty}\mathbb{E}\{x_i(k)-x_j(k)\}=0$ for all i and j.

(ii) The *expected* attribute divergence is obtained if $\limsup_{k\to\infty}\max_{i,j}|\mathbb{E}\{x_i(k)-x_j(k)\}|=\infty$.

2.3.1. Node pair selection

The actual interactions are selected using the following model: nodes interact with each other in the moment of a rate-one Poisson process and a node was selected randomly to interact with the others in each of these moments. Under this model, only one node or none originates an interaction at a given time. Ordering interaction events promptly and concentrating on modeling the node pair which is selected at interaction times. The node selection process is characterized by an $n\times n$ stochastic matrix $P=[p_{ij}]$, complying with the graph G in the sense that $p_{ij}>0$ always implies $\{i,j\}\in$ E for $i\neq j\in V$. We use p_{ij} to denote the probability that node i interacts with node j immediately after the node pair selection is executed as follows.