Pattern Recognition 43 (2010) 1334–1345

Contents lists available at ScienceDirect

Pattern Recognition

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

# Detecting the fuzzy clusters of complex networks

Jian Liu

*LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China*

ARTICLE INFO

*Article history:*
Received 24 March 2009
Received in revised form
22 September 2009
Accepted 9 November 2009

*Keywords:*
Fuzzy clusters
Optimal prediction
*k*-Means
Fuzzy *c*-means
Steepest descent
Conjugate gradient
Projection

ABSTRACT

To find the best partition of a large and complex network into a small number of clusters has been addressed in many different ways. However, the probabilistic setting in which each node has a certain probability of belonging to a certain cluster has been scarcely discussed. In this paper, a fuzzy partitioning formulation, which is extended from a deterministic framework for network partition based on the optimal prediction of a random walker Markovian dynamics, is derived to solve this problem. The algorithms are constructed to minimize the objective function under this framework. It is demonstrated by the simulation experiments that our algorithms can efficiently determine the probabilities with which a node belongs to different clusters during the learning process. Moreover, they are successfully applied to two real-world networks, including the social interactions between members of a karate club and the relationships of some books on American politics bought from Amazon.com.



## 1. Introduction

In recent years an explosive growth of interest and activity on the structure and dynamics of complex networks [1–3] has appeared. This is partly due to the influx of new ideas, particularly ideas from statistical mechanics, to the subject, and partly due to the emergence of interesting and challenging new examples of complex networks such as the internet and wireless communication networks. Network models have also become popular tools in social science, economics, the design of transportation and communication systems, banking systems, powergrid, etc., due to our increased capability of analyzing the models. On a related but different front, recent advances in computer vision and data mining have also relied heavily on the idea of viewing a data set or an image as a graph or a network, in order to extract information about the important features of the images or more generally, the data sets [4,5].

To give a coarse definition about the study of complex networks from the viewpoints of applied mathematics, it is about the research of dynamical systems on graphs. The graph structure may be fixed, or time-varying; the dynamical system may be deterministic, or stochastic. Since these networks are typically very complex, it is of great interest to see whether they can be reduced to much simpler systems. Such issues have been addressed before. In particular, much effort has gone into partitioning the network into a small number of clusters [4–14]. And in a broader aspect, it is also closely related to the model reduction theory of differential equations [15]. These proposals in the literature are constructed from different viewing angles, and their numerical performance applied to a benchmark model—the ad hoc network with 128 nodes and known community structures—are summarized in [16].

In a previous paper [12], a *k*-means approach is proposed to partition the networks based on optimal prediction theory proposed by Chorin and coworkers [17,18]. The basic idea is to associate the network with the random walker Markovian dynamics [19], then introduce a metric on the space of Markov chains (stochastic matrices), and optimally reduce the chain under this metric. The final minimization problem is solved by an analogy to the traditional *k*-means algorithm [20,21] in clustering analysis. This approach is motivated by the diffusion maps [11] and MNCut algorithms in imaging science [4].

The current paper is along the lines of extending the *k*-means type clustering techniques to the partitioning of networks. In statistical literature, a widely used generalization of *k*-means algorithm is the fuzzy *c*-means (FCM) algorithm [22,23]. In this framework, each node has a certain probability of belonging to a certain cluster, instead of assigning nodes to specific clusters, which is called fuzzy clustering in some papers [14]. This idea is quite valuable since usually it is not well separated for most of networks and the extending fuzzy partitioning framework seems extremely meaningful [13]. For the nodes lying in the transition domain between different clusters, the fuzzy partition will be more acceptable. To obtain the hard clustering result, one only needs to threshold the weights. But the fuzzy clustering presents more detailed information than the hard one, and it gives more reasonable explanations in some cases.

*E-mail address:* dugujian@pku.edu.cn

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.11.007

We constructed two algorithms—the steepest descent method with projection (SDP) and the reduced conjugate gradient method with projection (CGP)—from minimizing the objective function under the generalized framework in this paper. According to two choices of projection operators P1, P2, we obtain the formulations—SDP1, SDP2, CGP1, CGP2—which have been applied to two artificial networks, including the ad hoc network and the sample network generated from Gaussian mixture model, as well as two real-world networks, including the karate club network and the political books network. The proposed algorithms are easy to be implemented with reasonable computational effort and the final results do make sense in the considered models. It is demonstrated by these experiments that the algorithms can always perform successfully during the learning process and lead to a good clustering result.

The rest of the paper is organized as follows. In Section 2, the hard partitioning framework based on the optimal prediction in [12] is briefly introduced, and the corresponding fuzzy partitioning formulation is derived. The algorithms, SDP and CGP, are described in detail in Section 3. Several simulation and practical experiments are conducted in Section 4 to demonstrate the efficiency of the proposed algorithms. The numerical results and performance are typically compared. Finally, we conclude the paper in Section 5. All details of the derivation are left in the Appendix.

## 2. Framework for fuzzy clustering of networks

In [12], a new strategy for reducing the random walker Markovian dynamics based on optimal prediction theory [17,18] is proposed. Let $G(S,E)$ be a network with $N$ nodes and $M$ edges, where $S$ is the nodes set, $E=\{e(x,y)\}_{x,y\in S}$ is the weight matrix and $e(x,y)$ is the weight for the edge connecting the nodes $x$ and $y$. A simple example of the weight matrix is given by the adjacency matrix: $e(x,y)=0$ or 1, depending whether $x$ and $y$ are connected. We can relate this network to a discrete-time Markov chain with stochastic matrix $p$ with entries $p(x,y)$ given by

$$p(x,y)=\frac{e(x,y)}{d(x)},\quad d(x)=\sum_{z\in S}e(x,z), \tag{1}$$

where $d(x)$ is the degree of the node $x$ [11,19,24]. This Markov chain has stationary distribution

$$\mu(x)=\frac{d(x)}{\sum_{z\in S}d(z)} \tag{2}$$

and it satisfies the detailed balance condition

$$\mu(x)p(x,y)=\mu(y)p(y,x). \tag{3}$$

The basic idea in [12] is to introduce a metric for the stochastic matrix $p(x,y)$

$$\|p\|_\mu^2=\sum_{x,y\in S}\frac{\mu(x)}{\mu(y)}|p(x,y)|^2 \tag{4}$$

and find the reduced Markov chain $\tilde{p}$ by minimizing the distance $\|\tilde{p}-p\|_\mu$. For a given partition of $S$ as $S=\bigcup_{k=1}^K S_k$ with $S_k\cap S_l=\emptyset$ if $k\neq l$, let $\hat{p}_{kl}$ be the coarse grained transition probability from $S_k$ to $S_l$ on the state space $\mathbb{S}=\{S_1,\ldots,S_K\}$ which naturally satisfies

$$\hat{p}_{kl}\geq 0\quad\text{and}\quad\sum_{l=1}^K\hat{p}_{kl}=1. \tag{5}$$

This matrix can be naturally lifted to the space of stochastic matrices on the original state space $S$ via

$$\tilde{p}(x,y)=\sum_{k,l=1}^K\mathbf{1}_{S_k}(x)\hat{p}_{kl}\mu_l(y), \tag{6}$$

where $\mathbf{1}_{S_k}(x)=1$ if $x\in S_k$ and $\mathbf{1}_{S_k}(x)=0$ otherwise, and

$$\mu_k(x)=\frac{\mu(x)\mathbf{1}_{S_k}(x)}{\hat{\mu}_k},\quad\hat{\mu}_k=\sum_{z\in S_k}\mu(z). \tag{7}$$

Based upon this formulation, we can find the optimal $\hat{p}_{kl}$ for any fixed partition. With this optimal form $\hat{p}_{kl}$, we further search for the best partition $\{S_1,\ldots,S_K\}$ with the given number of clusters $K$ by minimizing the optimal prediction error. This is the theoretical basis for constructing the $k$-means algorithm for network partition in [12].

In the above formulation of hard clustering, each node belongs to only one cluster after the partition. This is often too restrictive for the reason that nodes at the boundary among clusters share commonalities with more than one cluster and play a role of transition in many diffusive networks. This motivates the extension of the optimal partition theory to a probabilistic setting [13]. Here we use the terminology hard clustering since we take indicator function $\mathbf{1}_{S_k}(x)$ in Eq. (6) when the node $x$ belongs to the $k$-th cluster. Now it is extended to the fuzzy clustering concept where each node may belong to different clusters with nonzero probabilities at the same time. We denote such probability function as $\rho_k(x)$ to represent the probability which the node $x$ belongs to the $k$-th cluster with. Naturally we need the assumption that

$$\rho_k(x)\geq 0\quad\text{and}\quad\sum_{k=1}^K\rho_k(x)=1 \tag{8}$$

for all $x\in S$.

Similar as before, we define the transition probability matrix of the induced Markov chain as

$$\tilde{p}(x,y)=\sum_{k,l=1}^K\rho_k(x)\hat{p}_{kl}\mu_l(y),\quad x,y\in S, \tag{9}$$

where

$$\mu_k(x)=\frac{\rho_k(x)\mu(x)}{\hat{\mu}_k}\quad\text{and}\quad\hat{\mu}_k=\sum_{z\in S}\rho_k(z)\mu(z). \tag{10}$$

The idea of lifting the size of stochastic matrices is similar as the hard clustering case and it expresses the perspective that the node $x$ transits to $y$ through different channels from cluster $S_k$ to cluster $S_l$ with their corresponding belonging probability and stay there in equilibrium state. It is not difficult to show that $\tilde{p}(x,y)$ is indeed a transition probability matrix and satisfies the detailed balance condition with respect to $\mu$

$$\mu(x)\tilde{p}(x,y)=\mu(y)\tilde{p}(y,x) \tag{11}$$

if $\hat{p}_{kl}$ satisfies the detailed balance condition with respect to $\hat{\mu}$, that is

$$\hat{\mu}_k\hat{p}_{kl}=\hat{\mu}_l\hat{p}_{lk}. \tag{12}$$

Given the number of the clusters $K$, we optimally reduce the random walker dynamics by considering the following minimization problem:

$$\min_{\rho_k(x),\ \hat{p}_{kl}}J=\|p-\tilde{p}\|_\mu^2 \tag{13}$$

where

$$J=\sum_{x,y\in S}\frac{\mu(x)}{\mu(y)}|p(x,y)-\tilde{p}(x,y)|^2=\sum_{x,y\in S}\mu(x)\mu(y)\cdot\left(\sum_{m,n=1}^K\rho_m(x)\rho_n(y)\frac{\hat{p}_{mn}}{\hat{\mu}_n}-\frac{p(x,y)}{\mu(y)}\right)^2, \tag{14}$$

subject to the constraints Eqs. (5) and (8).