Noname manuscript No.
(will be inserted by the editor)

Bipartite modularity maximization: complexity and exact methods

Alberto Costa · Pierre Hansen · Fabio Roda · Manuel Ruiz

Received: date / Accepted: date

Abstract Clustering has been a subject of intense study in the last decades. Its aim is to find subsets of vertices, called clusters, which are more likely to be joined pairwise by an edge than vertices in different clusters. A famous method to formalize this idea is the maximization of the modularity, which represents the fraction of edges within the clusters minus the expected fraction of such edges in a random graph with the same distribution of degrees. Modularity has been recently extended to bipartite graphs, which can be used to represent many situations (e.g., recommender systems between users and items, information retrieval with documents and terms, graphs showing relationships between authors and papers). This paper shows that maximizing the bipartite modularity is NP-hard. Moreover, some mathematical programming exact formulations are presented, as well as the column generation extension. Results for some instances of the literature are reported.

Keywords Bipartite graphs · Clustering · Modularity maximization

A. Costa
Singapore University of technology and Design, 138682 Singapore
E-mail: costa@sutd.edu.sg

P. Hansen
GERAD, HEC Montréal, 3000 Chemin de la Côte-Sainte-Catherine, H3T 2A7 Montréal, Canada
E-mail: pierre.hansen@gerad.ca

F. Roda
LIX, École Polytechnique, F-91128 Palaiseau, France
E-mail: roda@lix.polytechnique.fr

M. Ruiz
Artelys, 75002 Paris, France
E-mail: manuel.jean.ruiz@gmail.com

1 Introduction

Clustering on graphs is a very important topic, which has been studied under different points of view during these last years. Given a graph $G = (V, E)$, where V is the set of vertices and E is the set of edges joining pair of vertices, clustering aims at finding partitions of G, where a partition is a split of V into disjoint pairwise nonempty subsets of vertices covering V, called clusters, where the vertices belonging to the same cluster are more likely to be connected pairwise than to vertices belonging to other clusters. In order to formalize this idea, different approaches were proposed. For example, one could specify some conditions that must be satisfied by each cluster of the partition. In [4] the authors proposed the concepts of cluster in the *strong* and *weak* sense: the former refers to a cluster C where, for each vertex belonging to C, the number of its neighbors within C is larger than the number of its neighbors outside C. In the latter the sum, for all the vertices belonging to C, of the difference between the number of neighbors within C and the number of neighbors outside C is positive. Recently a modified version of the strong criterion, called *almost-strong*, has been proposed in [1]: for each degree 2 vertex v_i the strong condition forces this vertex and its 2 neighbors v_j and v_k to be in the same cluster, even if v_j and v_k are connected to two eterogeneous sets of vertices. The almost-strong criterion is the same as the strong one, exept for each degree 2 vertex v_i, where it is sufficient that at least one between v_j and v_k is in the same cluster as v_i. Other criterion have been proposed in [5, 2, 3].

Alternatively, one can define an objective function to maximize or minimize.

In this paper we consider the so called modularity maximization, introduced in.

2 Problem definition

3 Complexity

4 Mathematical programming models for bipartite modularity maximization

4.1 Fortet based formulation

4.2 Binary decomposition formulation

4.3 Clique partitioning based formulation

5 Column generation

6 Computational results

7 Conclusions

Acknowledgements A. C. has been partially supported by IDC Grant IDSF1200101OH.

References

1. Cafieri, S., Caporossi, G., Hansen, P., Perron, S., Costa, A.: Finding communities in networks in the strong and almost-strong sense. Physical Review E **85**(4), 046,113 (2012)
2. Hu, Y., Chen, H., Zhang, P., Li, M., Di, Z., Fan, Y.: Comparative definition of community and corresponding identifying algorithm. Physical Review E **78**(2), 026,121 (2008)
3. Luccio, F., Sami, M.: On the decomposition of networks in minimally interconnected subnetworks. IEEE Transactions on Circuit Theory **16**(2), 184 – 188 (1969)
4. Radicchi, F., Castellano, C., Cecconi, F., Loreto, V., Parisi, D.: Defining and identifying communities in networks. Proceedings of the National Academy of Sciences of the U.S.A. **101**(9), 2658–2663 (2004)
5. Wasserman, S., Faust, K.: Social Network Analysis. Methods and Applications. Cambridge University Pres, New York, USA (1994)