



Community Structures in Bipartite Networks: A Dual-Projection Approach

David Melamed*

Department of Sociology, University of South Carolina, Columbia, South Carolina, United States of America

Abstract

Identifying communities or clusters in networked systems has received much attention across the physical and social sciences. Most of this work focuses on single layer or one-mode networks, including social networks between people or hyperlinks between websites. Multilayer or multi-mode networks, such as affiliation networks linking people to organizations, receive much less attention in this literature. Common strategies for discovering the community structure of multi-mode networks identify the communities of each mode simultaneously. Here I show that this combined approach is ineffective at discovering community structures when there are an unequal number of communities between the modes of a multi-mode network. I propose a dual-projection alternative for detecting communities in multi-mode networks that overcomes this shortcoming. The evaluation of synthetic networks with known community structures reveals that the dual-projection approach outperforms the combined approach when there are a different number of communities in the various modes. At the same time, results show that the dual-projection approach is as effective as the combined strategy when the number of communities is the same between the modes.

Citation: Melamed D (2014) Community Structures in Bipartite Networks: A Dual-Projection Approach. PLoS ONE 9(5): e97823. doi:10.1371/journal.pone.0097823

Editor: Kimmo Eriksson, Mälardalen University, Sweden

Received: March 14, 2014; **Accepted:** April 23, 2014; **Published:** May 16, 2014

Copyright: © 2014 David Melamed. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. An R script, code, and random seeds are available on the author's website to replicate the results of the simulations reported in this paper (<https://sites.google.com/site/melamedpubsupplementalfiles/>).

Funding: The author received no specific funding for this work.

Competing Interests: The author has declared that no competing interests exist.

* E-mail: melamed@mailbox.sc.edu

Introduction

Discovering the community structure of bipartite networks often entails either examining the community structure of one of the modes of the network, or determining the community structure of both modes simultaneously using a ‘combined’ approach. The former has been shown to result in a loss of substantial structural information [1], which in turn results in a loss of analysts’ ability to uncover the function and topology of the network as a whole. The latter may be inefficient at recovering the community structure of both modes because relations between nodes in one mode are inferred based on indirect ties through the second mode, and the computations for nodes in one of the modes are affected by the nodes in the other mode. As such, the present paper advocates a dual-projection approach to uncovering the community structure of bipartite networks.

Some instances of bipartite graphs include affiliation networks which link people to committees [2], or Petri nets which are used in computer science designs of concurrent systems [3]. The dual-projection approach to community detection in such bipartite networks entails analyzing the community structure of each mode independently and then combining the community solutions in a manner that maximizes within-community ties [4–6]. For example, an affiliation network may be transformed into a person-to-persons network defined by shared committee membership, and a committee-to-committees network defined by shared members [2]. Then any of the many one-mode community detection algorithms [7–8] can be leveraged to determine the community structure of each network, and these community solutions can be combined to identify the overall community structure of the entire bipartite network. In particular, this method

should outperform the combined approach when the community structures differ between the modes of bipartite networks.

One situation in which the community structures differ between the modes of bipartite networks occurs when there are a different number of communities between the modes. This situation highlights the differences between the combined approach and the dual-projection approach. When using the dual-projection approach, community relevant computations for one mode are not contingent on the structure of the other mode in the same way that they are when using the combined approach. To validate this claim, this paper presents results from network simulations of bipartite networks with different community structures between the modes of the networks. Simulation results indeed support the contention that the dual-projection approach outperforms the combined approach. Below the two approaches are described in more detail. Then, the results from two simulations are reported. The paper concludes with some implications of this work and some potential extensions to consider in the future.

Communities in Bipartite Networks

The community structure of a bipartite network may be discovered by a variety of means [5–6,9–12]. As noted above, commonly the community structure of only one of the modes is analyzed, so the bipartite network is projected into a one-mode network. Suppose we have a binary bipartite network \mathbf{B} linking artists to teams [13]. The off-diagonal entries in the product $\mathbf{B}\mathbf{B}^T$ tells us how many teams are shared by each pair of artists, while the diagonal tells us the number of teams to which each artist belongs. This may be analyzed as an unweighted binary projection, or as a weighted projection. The weighted projection

approach has been shown to yield superior results [5], but analyzing only one of the modes results in a non-negligible loss of structural information [1].

On the other hand, the community structure of both modes may be of interest. In this case, the community structure of both modes may be discovered simultaneously [6,9] in a ‘combined’ analysis. One simple approach to doing so is to construct a block off-diagonal meta-matrix [14,15] which entails the bipartite graph and its transpose. Such a network has the following form

$$\begin{bmatrix} 0_{i \times i} & \mathbf{B}_{i \times j} \\ \mathbf{B}_{j \times i}^T & 0_{j \times j} \end{bmatrix}$$

where $0_{i \times j}$ is an all-zero matrix with i rows and j columns. This structure enables standard one-mode community detection algorithms to provide a solution for both of the modes of a bipartite network simultaneously.

An alternative to the combined approach is to apply the recent dual-projection approach [1]. To do so, take the weighted projection of both modes of a bipartite network, analyze the community structure of both modes separately, and then combine the community partitions in a fashion that maximizes within-community ties. Using the artist-to-teams network, for example, compute the community structure of artists using $\mathbf{B}\mathbf{B}^T$, then compute the community structure of teams using $\mathbf{B}^T\mathbf{B}$, and then maximize over one of the bipartite modularity extensions [5–6] to the standard one-mode modularity [4]. This approach also has the benefit of being amenable to all of the standard one-mode community detection algorithms, meaning that no new algorithms or software applications are required for its use. Below I discuss why this strategy is preferred to the combined approach that is discussed above.

The combined approach to identifying communities in networked systems has two related shortcomings. First, all information about the community structure pertaining to one of the modes is indirect. This follows from the simple fact that direct ties between nodes within one of the modes cannot occur within bipartite networks. Second, the community-related computations for the allocation of nodes in one of the modes are affected by the community-related computations for the nodes in the other mode, and this is particularly problematic when the community structure differs between modes. Using the classic Girvan and Newman betweenness-based community detection algorithm [16], for example, all of the geodesics for the nodes in one of the modes require indirect ties through nodes in the other mode when the community structure is discovered using the combined approach. The same is true of other methods such as spectral partitioning [17] or extremal optimization [18]. To validate use of the dual-projection approach relative to the combined approach, below the results of simulated networks with an unequal number of communities between the modes are analyzed using both approaches.

Test of the Method

To test the performance of the dual-projection approach relative to the combined approach I have applied both of them to a large set of synthetic graphs. An R script is available on the author’s website to replicate the results of the simulations reported in this paper (<http://sites.google.com/site/melamedpubssupplementalfiles/>). In both cases the community structure was analyzed by maximizing modularity using the walktrap algorithm [19], which places nodes into communities based on neighborhood similarity from short

random walks (of 4 steps) and has been shown to be particularly effective at recovering the community structure of large networks. The synthetic bipartite graphs had a density of .125 [16] and within-community ties occurred with a probability of .9. Varying the probability of within-community ties was found to make no substantive impacts on the results of the simulations. Further, the degree distribution for the nodes in the first mode was constrained to be fixed, but the degree of the nodes in the second was not.

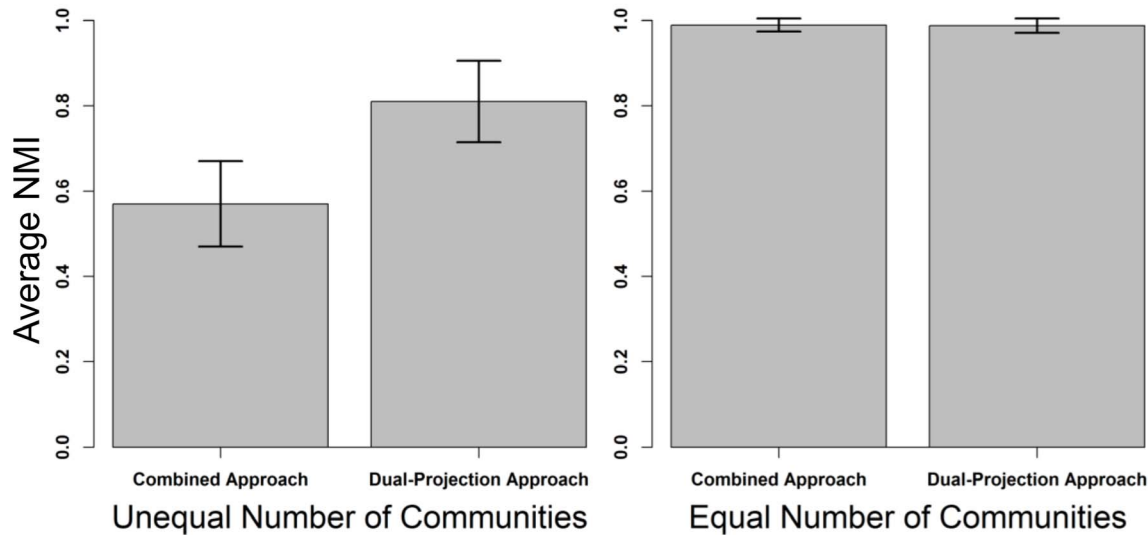
Two aspects of the synthetic graphs were manipulated. First, whether there was an unequal number of communities in the two modes was manipulated. Half of the graphs had three communities in the first mode and two in the second. To accomplish this, the first mode was divided into three equally-sized subsections. In the first subsection, within-community ties occurred within the first half of the nodes in the second mode. In the second subsection, within-community ties occurred within the middle third of the nodes in the second mode. Finally, in the third subsection, within-community ties occurred within the second half of the nodes in the second mode. This structure ensured that the most efficient allocation of nodes to non-overlapping communities entailed three communities in the first mode, and two in the second. The other half of the graphs had three equal-sized communities in both of the modes. This manipulation ensures that any increase in effectiveness for the dual-projection approach when there is an unequal number of communities does not come with a concomitant decrease in effectiveness when there is an equal number of communities in the two modes. Second, the size of the synthetic graphs was manipulated. Half of the networks were of dimensions 60×120 and the other half were of dimensions 600×1200 . This manipulation ensures that the results are robust to network size since many social networks are typically relatively small, while many physical networks are relatively large.

For the small networks, I relied on 1,000 realizations of each of the networks with equal and unequal community partitions. For the large networks, I relied on 100 realizations of the networks with equal and unequal community partitions. The large density of the large networks was set to .025. Each simulated network was analyzed using the combined and the dual-projection approach, and the normalized mutual information [20] between the known partition and the discovered partition was retained. Results rely on the average normalized mutual information across all of the simulated networks.

Figure 1a presents the results for the smaller networks. When the number of communities is unequal, it is evident that the dual-projection approach outperforms the combined approach (.81 and .57, respectively). At the same time, when the number of communities is equal, there is no difference in the effectiveness of the approaches (both .99). Figure 1b presents the results for the larger networks. When the number of communities is unequal, the discrepancy between the dual-projection approach and the combined approach gets larger: the dual-projection approach is clearly preferred (.97 and .22, respectively). Again, when the number of communities is equal, there is no difference between the approaches (both .99). These results suggest that the dual-projection approach should be preferred to the combined approach: it is more effective when the community structure differs between modes of a bipartite graph, but is just as effective when the community structure between the modes is the same.

While the above results show support for the use of the dual-projection approach when there are an unequal number of communities between the modes in a bipartite network, the two modes differed by only one community. To validate the method when the two modes have a greater disparity in the number of communities, I conducted a second simulation. Relying on the

A: 60 x 120 Networks



B: 600 x 1200 Networks

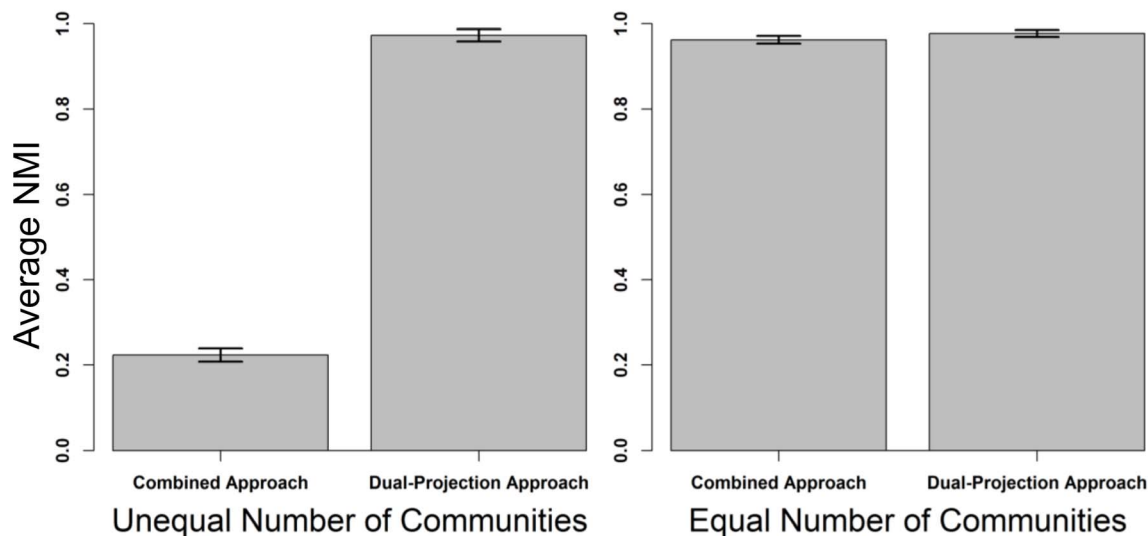


Figure 1. Average normalized mutual information between known and discovered community partitions. Each pair of graphs illustrates the average NMI for the combined approach and the dual-projection approach. The error bars refer to one standard deviation.
doi:10.1371/journal.pone.0097823.g001

same network size and probability of within-community ties, I generated networks with two communities in the first mode and ten in the second. To accomplish this, the nodes in the first mode were divided into twenty equally-sized subsections. In the first subsection, within-community ties occurred within the first tenth of the nodes in the second mode. In the second subsection, within-community ties occurred within the second tenth of the nodes in the second mode. The third subsection of nodes in the first mode linked the first two subsections by overlapping a thirtieth of the nodes in the second mode with the first and second tenths of the nodes in the second mode. Each of the remaining subsections in the first mode were allocated similarly, except that there was no overlap in the second mode between the tenth and eleventh subsections from the first mode, which ensured that a two community solution to the first mode was the optimal partition.

The simulated networks, as described above, were again analyzed using both the combined and dual-projection

approaches. One thousand small networks and 100 large networks were analyzed. Figure 2 presents the results. Again, the dual-projection approach outperforms the combined approach. The difference in the small networks is relatively small (.69 and .61, respectively), but the sample is quite large. As was observed in the first simulation, in the large networks the discrepancy becomes quite a bit larger (.85 and .45, respectively). Thus the dual-projection approach remains preferred with more unequal community structures between the modes of a bipartite network.

Conclusions

This paper illustrated that the dual-projection approach to detecting communities in bipartite networks should be preferred to the combined approach. In particular, sub-optimal community solutions may be identified when using the combined approach. As such, the results are more likely to misidentify sub-structures of

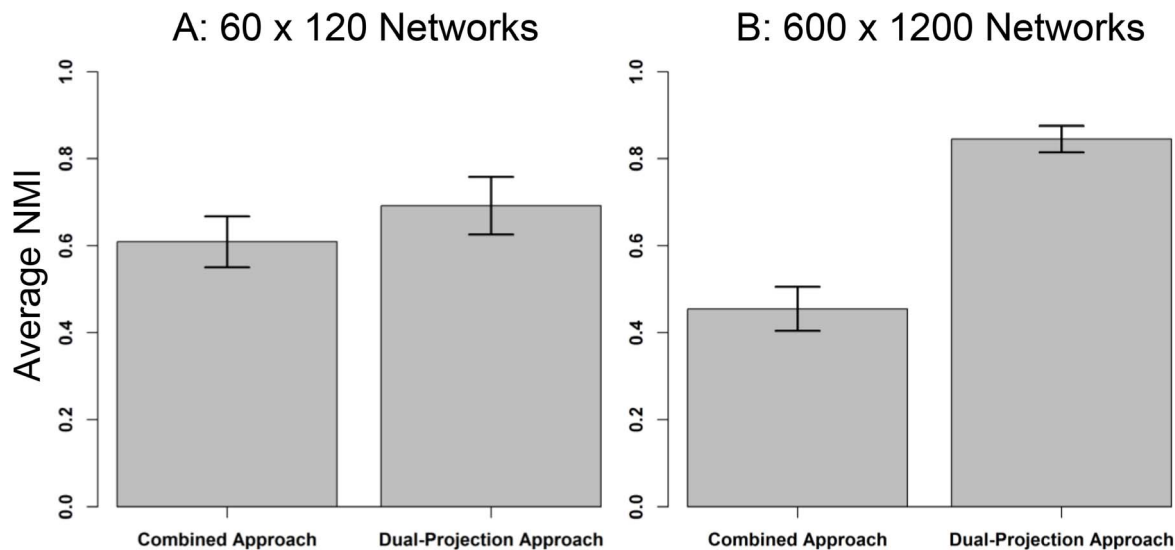


Figure 2. Average normalized mutual information between known and discovered community partitions. The pair of graphs illustrates the average NMI for the combined approach and the dual-projection approach. All networks had two communities in the first mode and ten in the second. The error bars refer to one standard deviation.
doi:10.1371/journal.pone.0097823.g002

bipartite networks, which in turn, may yield mistakes about the function and topology of the network. With respect to the dual-projection approach, all of the standard one-mode algorithms for community detection can be applied to discover the community structure of bipartite networks. That is to say, the increased accuracy of the dual-projection approach does not come with a corresponding loss of generality. Of particular relevance, the dual-projection approach has been shown to outperform the combined approach to detecting communities in bipartite networks when the community structure differs between modes of the network.

While it was assumed for the sake of simplicity up to this point that all of the nodes in graphs should be partitioned into unique communities, the dual-projection approach does not require this assumption. The dual-projection approach to partitioning bipartite graphs is particularly flexible and can be applied in conjunction with the evolving literatures on overlapping and hierarchical community structures [21–28]. Depending on the substance of the research question, it may be the case that one of the modes should be partitioned into non-overlapping communities, while the other mode should incorporate overlapping communities. For example, in a network of committee assignments it may be reasonable to assume that people may be allocated into overlapping communities while committees may not because they are often developed to account for discrete phenomenon. Likewise, one of the modes may entail hierarchical communities, while the other entails overlapping communities.

Another factor that has been ignored up to this point is that many networked systems have more than two modes. Some

literature deals with community structures in such networks [9,29]. The dual-projection approach can easily be generalized to k -partite graphs. The main insight here is that for any k , there are k choose two bipartite graphs that can be analyzed using the dual-projection approach. Of course, analysts can assume any form of community structure associated with the various modes embedded within the larger k -partite graphs.

In summary, it has been illustrated that the dual-projection approach to community detection in bipartite graphs is preferred to the combined approach. The dual-projection approach determines the community structure of each mode of bipartite networks separately while the combined approach determines them simultaneously. Results from the analysis of synthetic networks showed that the dual-projection approach outperforms the combined approach when the number of communities in the two modes of bipartite networks are unequal. At the same time, the dual-projection approach performs just as well when the number of communities in the two modes is equal. The dual-projection approach to community identification is flexible and warrants further investigation.

Author Contributions

Conceived and designed the experiments: DM. Performed the experiments: DM. Analyzed the data: DM. Contributed reagents/materials/analysis tools: DM. Contributed to the writing of the manuscript: DM.

References

- Everett MG, Borgatti SP (2013) The dual-projection approach for two-mode networks. *Social Networks* 35: 204–210.
- Breiger RL (1974) The duality of persons and groups. *Social forces* 53: 181–190.
- Jensen K, Kristensen LM, Wells L (2007) Coloured Petri Nets and CPN Tools for modelling and validation of concurrent systems. *International Journal on Software Tools for Technology Transfer* 9: 213–254.
- Newman ME, Girvan M (2004) Finding and evaluating community structure in networks. *Physical review E* 69: 026113.
- Guimerà R, Sales-Pardo M, Amaral LAN (2007) Module identification in bipartite and directed networks. *Physical Review E* 76: 036102.
- Barber MJ (2007) Modularity and community detection in bipartite networks. *Physical Review E* 76: 066102.
- Porter MA, Onnela J-P, Mucha PJ (2009) Communities in networks. *Notices of the AMS* 56: 1082–1097.
- Fortunato S (2010) Community detection in graphs. *Physics Reports* 486: 75–174.
- Melamed D, Breiger RL, West AJ (2013) Community structure in multi-mode networks: Applying an eigenspectrum approach. *Connections* 33: 18–23.

10. Zhang P, Wang J, Li X, Li M, Di Z, et al. (2008) Clustering coefficient and community structure of bipartite networks. *Physica A: Statistical Mechanics and its Applications* 387: 6869–6875.
11. Zhan W, Zhang Z, Guan J, Zhou S (2011) Evolutionary method for finding communities in bipartite networks. *Physical Review E* 83: 066120.
12. Freeman LC (2003) Finding social groups: A meta-analysis of the southern women data. *Dynamic social network modeling and analysis*: 39–97.
13. Guimerà R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308: 697–702.
14. Fararo TJ, Doreian P (1984) Tripartite structural analysis: Generalizing the Breiger-Wilson formalism. *Social Networks* 6: 141–175.
15. Carley KM (2003) Dynamic network analysis. *Dynamic social network modeling and analysis: Workshop summary and papers*. Citeseer. 133–145.
16. Girvan M, Newman ME (2002) Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99: 7821–7826.
17. Newman ME (2006) Modularity and community structure in networks. *Proceedings of the National Academy of Sciences* 103: 8577–8582.
18. Duch J, Arenas A (2005) Community detection in complex networks using extremal optimization. *Physical review E* 72: 027104.
19. Pons P, Latapy M (2006) Computing communities in large networks using random walks. *J Graph Algorithms Appl* 10: 191–218.
20. Strehl A, Ghosh J (2003) Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *The Journal of Machine Learning Research* 3: 583–617.
21. Reichardt J, Bornholdt S (2004) Detecting Fuzzy Community Structures in Complex Networks with a Potts Model. *Phys Rev Lett* 93: 218701. doi:10.1103/PhysRevLett.93.218701.
22. Xie J, Kelley S, Szymanski BK (2013) Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Computing Surveys (CSUR)* 45: 43.
23. Palla G, Derényi I, Farkas I, Vicsek T (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435: 814–818.
24. Lancichinetti A, Fortunato S, Kertész J (2009) Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics* 11: 033015.
25. Shen H, Cheng X, Cai K, Hu M-B (2009) Detect overlapping and hierarchical community structure in networks. *Physica A: Statistical Mechanics and its Applications* 388: 1706–1712.
26. Psorakis I, Roberts S, Ebdem M, Sheldon B (2011) Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E* 83: 066114.
27. Zhang S, Wang R-S, Zhang X-S (2007) Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A: Statistical Mechanics and its Applications* 374: 483–490.
28. Wang X, Jiao L, Wu J (2009) Adjusting from disjoint to overlapping community detection of complex networks. *Physica A: Statistical Mechanics and its Applications* 388: 5045–5056.
29. Murata T (2010) Detecting communities from tripartite networks. *Proceedings of the 19th international conference on World wide web*. ACM. 1159–1160.