Social Networking, **2014, 3, 80-85**
Published Online February 2014 (http://www.scirp.org/journal/sn)
http://dx.doi.org/10.4236/sn.2014.32010

Scientific Research

Clique Approach for Networks: Applications for Coauthorship Networks

Marcos Grilo Rosa[1], Inácio de Sousa Fadigas[1], Maria Teresinha Tamanini Andrade[2], Hernane Borges de Barros Pereira[3,4]

[1]Universidade Estadual de Feira de Santana, Feira de Santana, Brazil
[2]Instituto Federal de Educação Ciência e Tecnologia, Simões Filho, Brazil
[3]Universidade do Estado da Bahia, Salvador, Brazil
[4]Programa de Modelagem Computacional, SENAI Cimatec, Salvador, Brazil
Email: grilo@uefs.br, fadigas@uefs.br, tamanini@ifba.edu.br, hbbpereira@gmail.com

Received November 22, 2013; revised December 28, 2013; accepted February 6, 2014

ABSTRACT

Coauthorship networks consist of links among groups of mutually connected authors that form a clique. Classical approaches using Social Network Analysis indices do not account for this characteristic. We propose two new cohesion indices based on a clique approach, and we redefine the network density using an index of variance of density. We have applied these indices to two coauthorship networks, one comprising researchers that published in Mathematics Education journals and the other comprising researchers from a Computational Modeling Graduate Program. A contextualized and comparative analysis was performed to show the applicability and potential of the indices for analyzing social networks data.

KEYWORDS

Clique Networks; Cohesion Indices; Coauthorship

1. Introduction

Coauthorship networks are an example of social networks in which two authors are linked if they have written an article together. Due to the low number of authors, which is generally less than ten except for projects from large research groups, coauthorship networks represent a significant social involvement between authors. Because they comprise mutually connected groups (groups of an article's authors), these networks can be modeled using structures from graph theory known as *cliques*. Suppose there is a graph $G = (V, \varepsilon)$ with two sets: V, made of vertices (or nodes), and ε, consisting of elements called edges with one or two vertices connected to each [1]. A subgraph with a clique structure is the maximal subset of mutually adjacent vertices in G. For simple graphs (without loops or multiple edges), such as the graphs used in Social Network Analysis, a subgraph originating from a clique is a complete graph by definition.

Several aspects of coauthorship networks have been studied. For example, Katz and Martin [2] showed that using coauthorship to evaluate scientific collaboration is advantageous because, in addition to being invariant and verifiable, it is a relatively practical and inexpensive method. Similarly, although Vanz and Stump [3] distinguished between collaboration and coauthorship, designating coauthorship as one facet of scientific collaboration, this distinction has not hindered the use of coauthorship to assess collaboration, especially in bibliometrics and scientometrics. In the broader field of Social Network Analysis, where collaborations are treated as complex networks, Newman [4-6] studied the structure of scientific collaboration networks and found evidence for the small-world phenomenon. Maia and Caregnato [7] used degree centrality, betweenness centrality and closeness centrality to analyze a coauthorship network of professors in the Epidemiology graduate program at the Federal University of Pelotas (Universidade Federal de Pelotas—UFPel). Mello, Crubellate and Rossoni [8],

used density, number of components and centrality in coauthorship networks constructed using data extracted from the Lattes Platform to measure the level of collaboration in Administration graduate programs (*sensu stricto*). These studies used classical Social Network Analysis indices, especially centrality, to emphasize authors in the network.

On the other hand, there are some papers on clique and line graphs [9-11] that deal with the detection of communities or clusters. Evans [9] provided a method of community detection in networks based on line graphs. The proposed clique and line graphs are weighted graphs and of fixed order. Each vertex of the original graph is a clique and network analysis centers on the edges rather than the vertices as in the classical approach of Social Network Analysis. Evans applied their method on several networks of which we highlight co-authorship a network and a network of teams of a football league.

This article, however, seeks to redefine the classical concepts of density and show new cohesion indices for networks comprising cliques, as defined by Fadigas and Pereira [12], which will then be applied and interpreted using co-authorship networks. Thus, we intend to investigate relationships between cliques and not just relations between vertices (*i.e.* actors). The main cohesion index used to characterize a network, *i.e.*, density, is parameterized by two extreme cases of network topology: a network containing only vertices (density of 0) and a network where all the vertices are mutually connected (*clique*), *i.e.*, a network with a density of 1. In networks such as the ones involved in coauthorship, a topology with a null density does not exist by definition (authors who publish alone do not form coauthorships). At the other extreme, it would be difficult for real coauthorship networks to exhibit a density of one, as it would imply that all of the authors have published in collaboration with all of the other authors in the network at least once.

2. Clique Approach for Networks

In this section, we present an original approach for analyzing networks whose basic elements are *cliques*, proposing new cohesion indices and redefining classical indices [12]. The approach is based on the notion that initially isolated cliques form networks by *juxtaposition* and/or *superposition*. Fadigas and Pereira [12] define *juxtaposition* as the process where two cliques are linked by a single common vertex. The authors denote processes where cliques are linked by two or more common vertices by *superposition*.

Figure 1 shows an *initial configuration of disconnected cliques*, before the *juxtaposition* and/or *superposition* process. **Figure 2** shows an example of a network formed by applying the processes of juxtaposition and/or superposition to the initial configuration shown in **Figure 1**. The networks that result from these processes are called *clique networks*, in reference to the basic components that may or may not be connected.

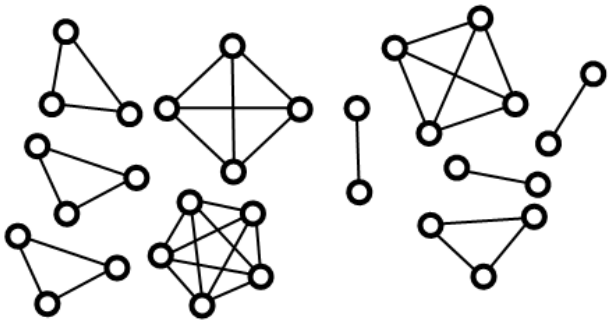


Figure 1. Initial configuration of disconnected cliques.

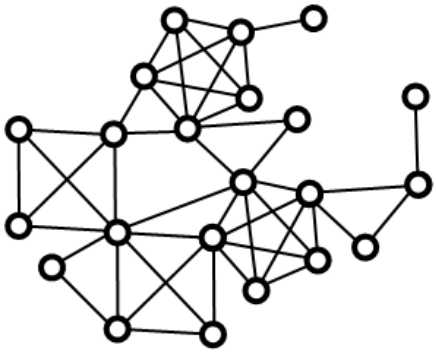


Figure 2. Formation of a *clique* network by *juxtaposition* and *superposition*.

3. Cohesion Index Based on a Clique Approach

From the initial configuration of disconnected cliques and the processes of *juxtaposition* and/or *superposition* that create the clique network, Fadigas and Pereira [12] propose two novel cohesion indices: normalized density and variance of density. In this article, three additional cohesion indices are introduced: edge superposition, vertex reduction factor and component reduction factor.

3.1. Normalized Density and Variance of Density

One of the main cohesion indices for Social Network Analysis is the density (Δ) for undirected networks with *n* vertices, which relates the number of edges in the network $(|\varepsilon|)$ to the maximum possible number of edges, given by $(n(n-1)/2)$. The *density* (Δ) of a network is an index that varies from 0 to 1. When $\Delta = 0$, the network is totally disconnected and does not adequately reflect the initial configuration of disconnected cliques (**Figure 1**). The density of the initial configuration of disconnected cliques can be calculated using the following expression:

$$\Delta_{q0} = \frac{2|\varepsilon|}{n_0(n_0 - 1)} \quad (1)$$

where n_0 is the number of vertices in the initial configuration of disconnected cliques. Fadigas and Pereira [12] proposed a more appropriate normalization for density in clique networks, denoted as the *normalized density* (Δ_{norm}):

$$\Delta_{norm} = \frac{\Delta - \Delta_{q0}}{1 - \Delta_{q0}} \quad (2)$$

when $\Delta_{norm} = 0$, the actual network is equivalent to the initial configuration of disconnected *cliques*. Fadigas and Pereira [12] also proposed the variance of density $(v(\Delta))$, which measures the densification of the network compared to the initial configuration of disconnected cliques and can be calculated by the following expression:

$$v(\Delta) \cong \frac{|\varepsilon|}{|\varepsilon_0|} \times \frac{n_0^2}{n^2} - 1 = \frac{[n_0/n]^2}{[|\varepsilon_0|/|\varepsilon|]} - 1 \quad (3)$$

3.2. Edge Superposition

The *variance of density* depends on both the relationship between the edges and the relationship between the vertices. Therefore, it does not directly quantify *clique* superposition. R_ε is a superposition rate that compares the number of edges in the initial configuration of disconnected *cliques* with the number of edges after juxtaposition and/or superposition. This rate is defined by Equation (4) and is applicable to both connected and disconnected networks.

$$R_\varepsilon = \frac{|\varepsilon_0| - |\varepsilon|}{|\varepsilon_0| - \varepsilon_{\max}} \quad (4)$$

In the above equation, $\varepsilon_{\max}$ is the number of edges in the largest clique present in the initial configuration of disconnected cliques. A superposition value of 0 occurs when $\varepsilon = \varepsilon_0$ and, therefore, there is no superposition of the edges, although juxtaposition may (or may not) occur. A superposition value of 1 occurs when $\varepsilon = \varepsilon_{\max}$, *i.e.*, when the network comprises only the largest clique.

3.3. Vertex Reduction Factor

In parallel to R_ε for the superposition of edges, (R_V) is a factor that measures the vertex juxtaposition rate and can be used to compare the number of vertices in the initial configuration of disconnected cliques with the number of vertices in the network resulting from juxtaposition and/or superposition. The difference between the number of vertices in the initial configuration of disconnected *cliques* and the number of vertices resulting from the juxtaposition and/or superposition processes shows how many vertices are shared between the two cliques, and it can be parameterized by the number of vertices in the resulting network. Symbolically, again denoting the number of vertices in the initial configuration of disconnected *cliques* by n_0, the number of vertices in the largest *clique* of the initial configuration of disconnected *cliques* by $n_{\max}$ and the number of vertices in the network after the juxtaposition and/or superposition processes by n, we obtain

$$R_v = \frac{n_0 - n}{n_0 - n_{\max}} \quad (5)$$

The index varies from 0 to 1. A null value occurs when $n = n_0$, *i.e.*, when there is no juxtaposition or superposition. The maximum value, 1, occurs when the network comprises only the largest clique in the initial configuration.

3.4. Component Reduction Factor

We observed in Section 3.1 that the number of cliques is represented by n_q in the initial clique configuration. We consider this number to be the number of "components" in the initial configuration. Thus, it is also possible to quantify the reduction in the number of components as a measure of the network cohesion. Therefore, the component reduction (R_C) measures how often *cliques* from the initial configuration are connected to form larger components, consequently reducing their number. It is normalized in the same way as the previous indices, using the minimum number of components that can result as the base, which is 1. Considering that n_q can be defined as the number of components in the initial configuration of disconnected *cliques*, let C_C be the number of components in the network. Thus, the factor can be expressed as follows

$$R_C = \frac{n_q - C_C}{n_q - 1} \quad (6)$$

The values for R_C vary from 0 to 1. A value of 0 occurs when there is no juxtaposition or superposition. Conversely, a value of 1 results when the network is connected.

4. Application to Coauthorship Networks

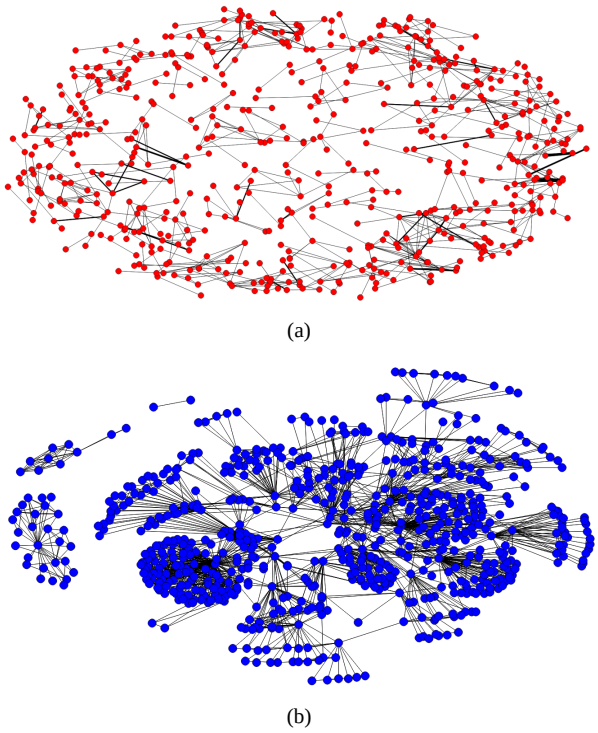
After defining the indices using this new approach, we chose two coauthorship networks to calculate and interpret the indices. One of the networks consists of authors who published in six Mathematics journals, which we grouped under the name of Mathematics Education journals. The other network consists of researchers in a computational modeling graduate program.

The coauthorship networks were constructed so that each group of coauthors is mutually connected, and each

group is connected to another if they have at least one author in common. To interpret coauthorship as a type of collaboration, we excluded from the network authors who only published alone. We should note that the ME and GP networks have 1000 and 795 vertices and 572 and 11 components, respectively, before authors who published alone were excluded. **Figures 3(a)** and **(b)** show the networks, and **Table 1** summarizes some properties for the networks.

5. Results and Interpretation

As the indices from our *clique* approach were applied to only two networks, we performed an interpretive analysis of the new indices proposed here and compared the two. **Table 1** shows that the ME and GP networks possess numbers of vertices of the same order of magnitude (588 and 792, respectively). The same is true for the number of coauthored publications (944 and 1030, respectively). However, the number of authors in the GP Network, represented by the number of vertices in the initial configuration of the disconnected cliques, is almost twice the number of authors in the ME Network. One significant difference between the networks is the relative size of the largest component, *i.e.*, the largest group of interconnected authors in the network. For the ME Network, the largest group contains only approximately 7% of the authors; however, this number is approximately 90% in the GP Network. This difference is due to the distinct nature of the two networks: while the ME Network contains authors who publish in six distinct journals, the GP Network comprises only researchers who are part of the same graduate program. However, despite the distinct nature of the Mathematics Education journals, they form the *corpus* of publications on the field, and the relatively small number of authors in a single group allows little collaboration, in terms of coauthorship.



(a)

(b)

Figure 3. Coauthorship networks: (a) ME Network, where the thickness of the edges is proportional to the number of studies published by the corresponding pair of vertices; (b) GP Network, with the thickness of the edges calculated as described above. (a) ME Network; (b) GP Network.

Table 1. Basic properties of the ME and GP networks.

Properties	ME network	GP network
Number of vertices (n)	588	792
Number of authors	1751	3234
Author/vertex ratio	2.9779	4.083
Components	160	8
Largest component (%)	7.14	90.28
Coauthored publications	944	1030
Mean *clique* size	1.85	2.97

The cohesion indices calculated for the two networks are shown in **Table 2**. The density (binary) is calculated without accounting for the number of times that a pair of authors published together, while the valued density does account for repeated joint publications.

The ratio between the two numbers provides the mean number of publications per pair and reflects how often authors publish together. **Table 2** shows that, in the ME Network, each pair published once, on average, while the mean is nearly two in the GP Network.

The variance of density measures the "densification" of the network, *i.e.*, how much the groups of coauthors coalesce, either in isolation or in pairs, when the network is formed. The index measures the variance of the vertices and the edges simultaneously, compared to the initial

Table 2. Cohesion indices using a clique approach.

Properties	ME network	GP network
Density (binary) (Δ)	0.0037	0.0073
Density (valued) (Δ_{valued})	0.0040	0.0135
Δ/Δ_{valued}	1.1029	1.8554
Normalized density (Δ_{norm})	0.0036	0.0127
Variance of density $(v(\Delta))$	7.8784	15.4440
Edge superposition (R_e)	0.0972	0.5189
Vertex reduction factor (R_v)	0.6672	0.7857
Component reduction factor (R_C)	0.8736	0.9936

configuration of disconnected cliques. The results in **Table 2** show that the GP Network displays a higher variance of density index score higher than the ME Network. The theoretical maximum value for the variance of density occurs when the network becomes a single clique, *i.e.*, all of the vertices are mutually connected. For the ME Network, this maximum value is approximately 2200, and thus the value found herein (7.8784) corresponds to only 0.4% of the maximum. In the case of the GP Network, the maximum variance of the density is 1223, and the measured value (15.4440) is 1.3% of the maximum. Comparatively, it can be stated that the GP Network is more than threefold "densified" compared to the ME Network. The variance of density index reflects the coalescence of the authors that published alone but also as coauthors (reduction of vertices without a reduction in the number of edges) in relation to the coalescence of pairs that published as coauthors (simultaneous reduction of vertices and edges). To more precisely determine what type of situation predominates in the network, the superposition of edges and the vertex reduction factor can be used.

The superposition of edges determines the proportion of coauthor pairs in common; *i.e.*, it is an index that measures relationships, represented by the edges in the network. The superposition of approximately 10% in the ME Network indicates that few pairs of authors published together more than once, unlike the GP Network, where 52% of the pairs have more than one publication together. This index, therefore, shows that there is more scientific production by pairs of authors in the GP Network than in the ME Network.

The vertex reduction factor, in turn, is directly related to the author/vertex ratio in **Table 1**. This index indicates the percentage of authors with more than one publication. The values obtained for the ME and GP networks indicate that the researchers in the GP network display higher individual productivity.

The values of the component reduction factor for the two networks do not differ. Although the ME Network has 160 components compared to only 08 for the GP Network, the percentage difference is approximately 12%. However, almost all of the coauthorship groups shared at least one vertex in common (link) in the GP Network (99.4%), while a value of 87.4% was observed in the ME Network.

6. Final Considerations

The clique approach in coauthorship networks allows the social data to be analyzed in a way that is well suited for the network topological structure. Network analysis using cohesion indices already allows new interpretations. For example, considering the index that measures edge superposition together with the vertex reduction factor allowing us to clarify how the juxtaposition and superposition processes create the network. Thus, we observed that superposition predominated in the GP Network compared to the ME Network. This effect also occurred with the vertex reduction factor, but to a lesser extent. These aspects result in a greater "densification" of the GP Network, mostly due to the large number of pairs of authors who have written more than one study together. These results are consistent with the fact that the GP Network comprises researchers connected through the same research institution, while the ME Network includes researchers who may have stronger ties within their own groups, but this collaboration is not shown through their publication in journals of the field.

The initial research using cohesion indices showed that other indices could potentially be added, and the dynamics of network growth could be evaluated. Another aspect that we emphasize is the applicability of the clique approach to other social networks with a similar structure, such as actor-movie networks.

Finally, it is important to comment that this work is an ongoing research and initially it was published in the proceedings of the 1st Brazilian Workshop on Social Network Analysis and Mining [13].

REFERENCES

[1] J. L. Gross and J. Yellen, "Graph Theory and Its Applications. Discret Mathematics and Its Applications," CRC Press, Boca Raton, 2003.

[2] J. S. Katz and B. R. Martin, "What Is Research Collaboration?" *Research Policy*, Vol. 26, No. 1, 1997, pp. 1-18. http://dx.doi.org/10.1016/S0048-7333(96)00917-1

[3] S. A. d. S. Vanz and I. R. C. Stump, "Scientific Collaboration: A Theoretical-Conceptual Review," *Perspectivas em Ciência da Informação*, Vol. 15, No. 2, 2010, pp. 42-55. http://dx.doi.org/10.1590/S1413-99362010000200004

[4] M. E. J. Newman, "The Structure of Scientific Collaboration Networks," *Proceedings of the National Academy of Sciences*, Vol. 98, No. 2, 2001, pp. 404-409. http://dx.doi.org/10.1073/pnas.98.2.404

[5] M. E. J. Newman, "Scientific Collaboration Networks. I. Network Construction and Fundamental Results," *Physical Review E*, Vol. 64, No. 1, 2001, Article ID: 016131.

[6] M. E. J. Newman, "Scientific Collaboration Networks. II. Shortest Paths, Weighted Networks, and Centrality," *Physical Review E*, Vol. 64, No. 1, 2001, Article ID: 016132.

[7] M. F. Maia and S. E. Caregnato, "Co-Authorship as an Indicator of Scientific Collaboration Network," *Perspectivas em Ciência da Informação*, Vol. 13, No. 2, 2008, pp. 18-31.

[8] C. M. Mello, J. M. Crubellate and L. Rossoni, "Coauthor Networks between Brazilian Graduate Administration Program Faculty (Strictu Sensu): Structural and Dynamic Aspects of Relationships," *Perspectivas em Ciência da Informação*, Vol. 15, No. 2, 2009, pp. 42-55.

[9] T. S. Evans, "Clique Graphs and Overlapping Communities," *Journal of Statistical Mechanics*: *Theory and Experiment*, Vol. 2010, 2010, Article ID: P12037. http://dx.doi.org/10.1088/1742-5468/2010/12/P12037

[10] T. S. Evans and R. Lambiotte, "Line Graphs of Weighted Networks for Overlapping Communities," *The European Physical Journal B*, Vol. 77, No. 2, 2010, pp. 265-272. http://dx.doi.org/10.1140/epjb/e2010-00261-8

[11] T. S. Evans and R. Lambiotte, "Line Graphs, Link Partitions, and Overlapping Communities," *Physical Review E*, Vol. 80, No. 1, 2009, Article ID: 016105. http://dx.doi.org/10.1103/PhysRevE.80.016105

[12] I. S. Fadigas and H. B. B. Pereira, "A Network Approach Based on Cliques," *Physica A*: *Statistical Mechanics and its Applications*, Vol. 392, No. 10, 2013, pp. 2576-2587.

[13] M. G. Rosa, I. S. Fadigas, M. T. T. Andrade and H. B. B. Pereira, "Clique Approach for Networks: Applications for Coauthorship Networks," *Proceedings of the Brazilian Workshop on Social Network Analysis and Mining, XXXII Congress of the Brazilian Computer Society Computer Society*, Curitiba, 2012.