Conducting a bipartite network analysis to assess people's behavioural patterns

Ismagilova Adelina0000-0002-6596-5861

[1]ITMO University

This manuscript was compiled on January 14, 2021

This paper proposes a bipartite network analysis based on graph theory. The estimation of the behavioural model of a group of people will be done on the basis of bipartite graphs, and the centralisation of values, measures and their interrelationships will be calculated on their basis.

bipartite network model | graph theory | behavioural model

Evaluation of human behaviour patterns is widespread in both scientific and industrial fields [1]. One of the most popular methods for conducting analysis is the use of graphs. Thanks to graphs, it is possible to consider and evaluate the relationships between users in order to propose possible collaborations, as well as to perform crowd control at large events. Thus, through a bivariate network analysis, it will be possible to determine the following format of information: e.g. which banks people use, which events people attend and what is the relationship between the visitors.

The most relevant method for conducting analysis is the subject-oriented approach for counting the links between the user and the organisation. However, simple methods cannot fully capture the systemic nature of the process, as they usually analyse information from the perspective of only one actor, resulting in one-dimensional dimensions of need. Thus, there is a need to create a more complex method that allows grouping of vertices, namely the bipartite network method. [2]

Bipartite networks are a special kind of networks that are crucial for the analysis of social and economic systems because they explicitly represent conceptual relationships between different types of actors. They can be used to show relationships based on membership, affiliation, cooperation, employment, ownership and other relationships between two classes of nodes within a system, namely lower and upper nodes. [3,4]

This paper explores graph theory, analyses human behaviour and identifies bipartite networks, and uses them to model group relationships.

Methodology

Social network analysis is an analytical approach to the study of social structures and phenomena based on graph theory. The approach is relational and structural in that it focuses on the patterns of relationships between objects in a social system, modelling them as network structures. Such network structures are characterised as nodes (individual actors, people) and boundaries (relationships or connections) that link them. The individual actors and people are represented as nodes in the network, and the relationships between them as lines that link them. Such visualisation can help to provide a more qualitative and quantitative assessment of the networks upon which human behavioural patterns are analysed. These structures are commonly drawn as graphs. [5,6]

Graphs consist of a set of nodes $V(G)$ and a set of nodes $E(G)$ from the number of nodes m , each of which connects a pair of nodes, and are described by the formula $G = (V, E)$. The order G is the number of vertices n and the size G is the number of edges m Vertices will be considered adjacent if they are connected by edges that randomly connect two vertices. Let us also consider the most important terms. [7]

Terms and Definitions. A subgraph G is a graph with all vertices and edges in G A called subgraph $G[U]$ is a subgraph defined by a subset of vertices in $U \subseteq\ V(G)$ with all edges in Glinking vertices in U A subgraph is maximal with respect to some property if adding more nodes to the subgraph will lead to loss of that property.

A path is an alternating sequence of individual vertices and edges in which each edge falls with the preceding and succeeding vertices. The length of a path is the number of edges it contains. A graph is connected if each pair of vertices is connected by at least one path. The shortest path between two vertices is the path with the minimum number of edges. The distance between any two vertices u and v G denoted by $dG(u, v)$ is the length of the shortest path between them. The diameter of the graph G is the length of the longest shortest path between any pairs of nodes G.[8]

The graph density G denoted by $\varrho(G)$, measures how many edges in the set $E(G)$ compared to the maximum possible number of edges between vertices in $V(G)$. Thus the density is calculated as $\varrho(G) = 2mn(n-1)$. [9]

The degree of a vertex v denoted by $deg(v)$ is the number of edges that fall on v. The minimal degree of graph G is denoted by $\delta(G)$ and is the smallest degree of a vertex in G.

As described earlier, a bipartite network must be used to perform the analysis. A bipartite network is a graph $B = U, V, E$, where U and V are disjoint. sets of nodes and $E = (u, v) : u \in U, v \in V$ is a set of linking nodes in different sets. Each set of nodes in a bipartite network can have independent properties such as degree distribution, clustering, number of nodes (system size). Let us refer to sets U and V as lower and

Significance Statement

It is now becoming increasingly relevant to account for the number of users of a particular product, the behaviour of each person at an event and the possible correlation between them. The relevance lies in the fact that bi-directional networks will make it easier for organisations to identify the customer market and find new potential customers, as well as opening up more opportunities for research into human behaviour.

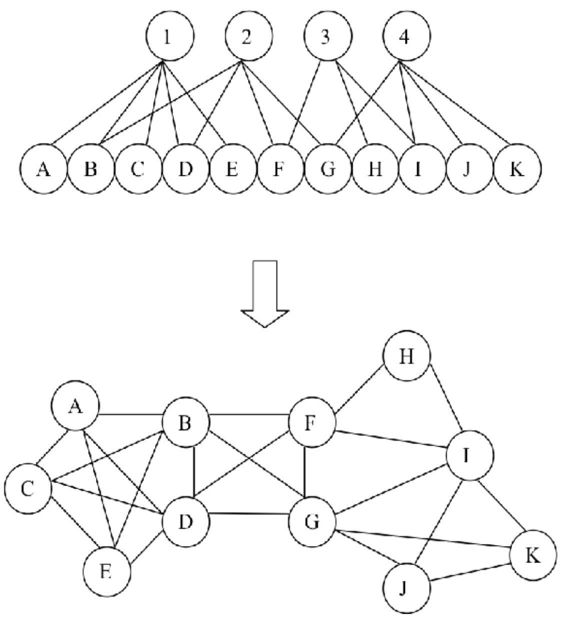

Fig. 1. An example of unipartite and bipartite graphs.

upper sets respectively. Nodes in U can only connect to nodes in V and vice versa. Hence, in bipartite networks there are

- k_u: the degree of a node $u \in U$;

- d_v: the degree of a node $v \in V$;

- $P_b(k)$: degree of distribution of lower vertices in U;

- $P_t(d)$: the degree of distribution of the top vertices in V.

Projection onto nodes U leads to a single-mode network G, where node u is connected to u‘ with u, u‘ $\in U$, only if there is a pair of edges (u, v) and $(u‘, v)$ in E.

-g_u: degree of node $u \in U$, in G; -$P(q)$: degree distribution for nodes $u \in U$. [10]

In order to assess the pattern of people’s behaviour, a centralisation analysis of the nodes (actors) must be carried out, to determine their measure and interconnectedness. The centralisation analysis identifies the most important nodes (i.e. actors) of the network. Determining the importance depends on the network (event, organisation) being analysed and the nature of the relationships it models. Thus, it is necessary to determine:

- The degree of a vertex in an undirected graph is the number of incident edges of the vertex, i.e. the number of neighbours it has in the graph. The degree of mesor centrality is the number of neighbours of each vertex divided by the maximum number of neighbours it can have. In an undirected graph it is $n - 1$, where n is the total number of vertices of the graph.

- Betweenness centrality of a node v is the sum of the fraction of all-pairs shortest paths that pass through v:

$$c_B\ (v) = \sum_{(s,t \in V)} \frac{\sigma(s,t|v)}{\sigma(s,t)} \qquad [1]$$

where V is the set of nodes, $\sigma(s, t)$ is the number of shortest (s, t)-paths, and $\sigma(s, t|v)$ is the number of those paths passing through v. - Closeness centrality of a node u is the reciprocal of the sum of the shortest path distances from u to all n-1 other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalized by the sum of minimum possible distances $n - 1$.

$$C(u) = \frac{n-1}{\sum_{v=1}^{n-1} d(v,u)} \qquad [2]$$

where $d(v, u)$ is the shortest-path distance between v and u, and n is the number of nodes in the graph. Notice that higher values of closeness indicate higher centrality.[11]

Results and Discussion

As an example, let us define a grouping of links from a dataset of girls who attended 14 informal social events. [12].

Allison Davis et. al was concerned with the issue of how much the informal contacts made by individuals were established solely (or primarily) with others at approximately their own class levels. To address this question the authors collected data on social events and examined people’s patterns of informal contacts.

In particular, they collected systematic data on the social activities of 18 women whom they observed over a nine-month period. During that period, various subsets of these women had met in a series of 14 informal social events. The participation of women in events was uncovered using "interviews, the records of participant observers, guest lists, and the newspapers", who presumably had been in touch with the research team, reported that the data reflect joint activities like, "a day’s work behind the counter of a store, a meeting of a women’s club, a church supper, a card party, a supper party, a meeting of the Parent-Teacher Association, etc."

This data set has several interesting properties. It is small and manageable. It embodies a relatively simple structural pattern, one in which, according to Davis et al., the women seemed to organize themselves into two more or less distinct groups. Moreover, they reported that the positions - core and peripheral - of the members of these groups could also be determined in terms of the ways in which different women had been involved in group activities.

At the same time, the Souther Women data set is complicated enough that some of the details of its patterning are less than obvious. Thus, it provides an opportunity to explore methods designed for direct application to two-mode data. But at the same time, it can easily be transformed into two one-mode matrices (woman by woman or event by event) that can be examined using tools for one-mode analysis. [13,14]

Initially, let’s define the nodes of the graph. The nodes of the two-sided graph will be both women and events, as each woman is associated with events she attended.

As a result of the data, a graph was drawn to more fully depict the relationship between women and the activities they attended (Fig.2).

The process of obtaining an unipartite graph with only women that are linked if they attended the same events is called projection. Let’s try to weight the edges of the projection using different criteria. For example, let’s determine the weight of an edge to reflect the number of events each woman participated in. Weighting was done using NetworkX and the result is the following graph, which clearly shows which women were in contact with which women when they attended the same events (Fig.3).

The centralisation of values, measures and their relationships were then determined. In order to compute the cen-

```
[('Evelyn Jefferson', {'bipartite': 0}),
 ('Laura Mandeville', {'bipartite': 0}),
 ('Theresa Anderson', {'bipartite': 0}),
 ('Brenda Rogers', {'bipartite': 0}),
 ('Charlotte McDowd', {'bipartite': 0}),
 ('Frances Anderson', {'bipartite': 0}),
 ('Eleanor Nye', {'bipartite': 0}),
 ('Pearl Oglethorpe', {'bipartite': 0}),
 ('Ruth DeSand', {'bipartite': 0}),
 ('Verne Sanderson', {'bipartite': 0}),
 ('Myra Liddel', {'bipartite': 0}),
 ('Katherina Rogers', {'bipartite': 0}),
 ('Sylvia Avondale', {'bipartite': 0}),
 ('Nora Fayette', {'bipartite': 0}),
 ('Helen Lloyd', {'bipartite': 0}),
 ('Dorothy Murchison', {'bipartite': 0}),
 ('Olivia Carleton', {'bipartite': 0}),
 ('Flora Price', {'bipartite': 0}),
 ('E1', {'bipartite': 1}),
 ('E2', {'bipartite': 1}),
 ('E3', {'bipartite': 1}),
 ('E4', {'bipartite': 1}),
 ('E5', {'bipartite': 1}),
 ('E6', {'bipartite': 1}),
 ('E7', {'bipartite': 1}),
 ('E8', {'bipartite': 1}),
 ('E9', {'bipartite': 1}),
 ('E10', {'bipartite': 1}),
 ('E11', {'bipartite': 1}),
 ('E12', {'bipartite': 1}),
 ('E13', {'bipartite': 1}),
 ('E14', {'bipartite': 1})]
```

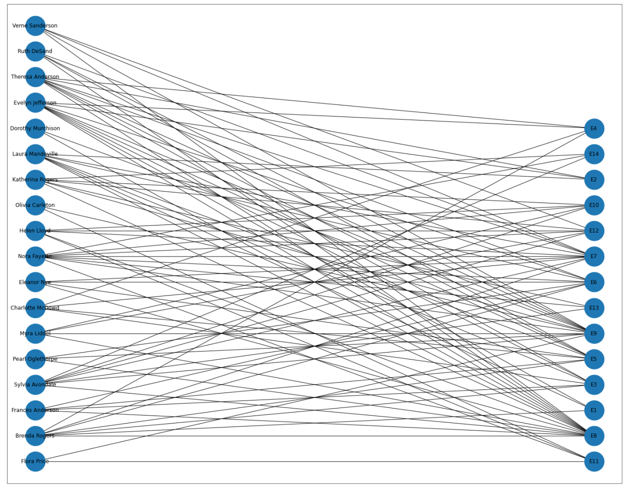

Fig. 2. Bipartite graph showing the relationship between women and activities.

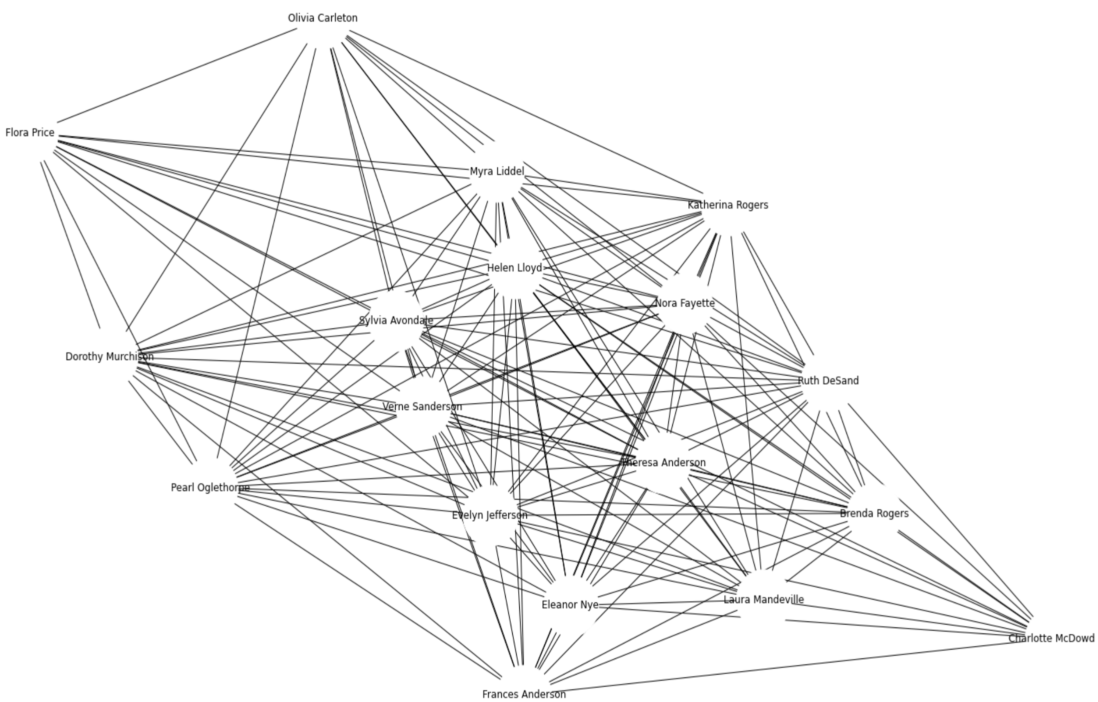

Fig. 3. A graph of the relationships between the women Allison Davis studied.

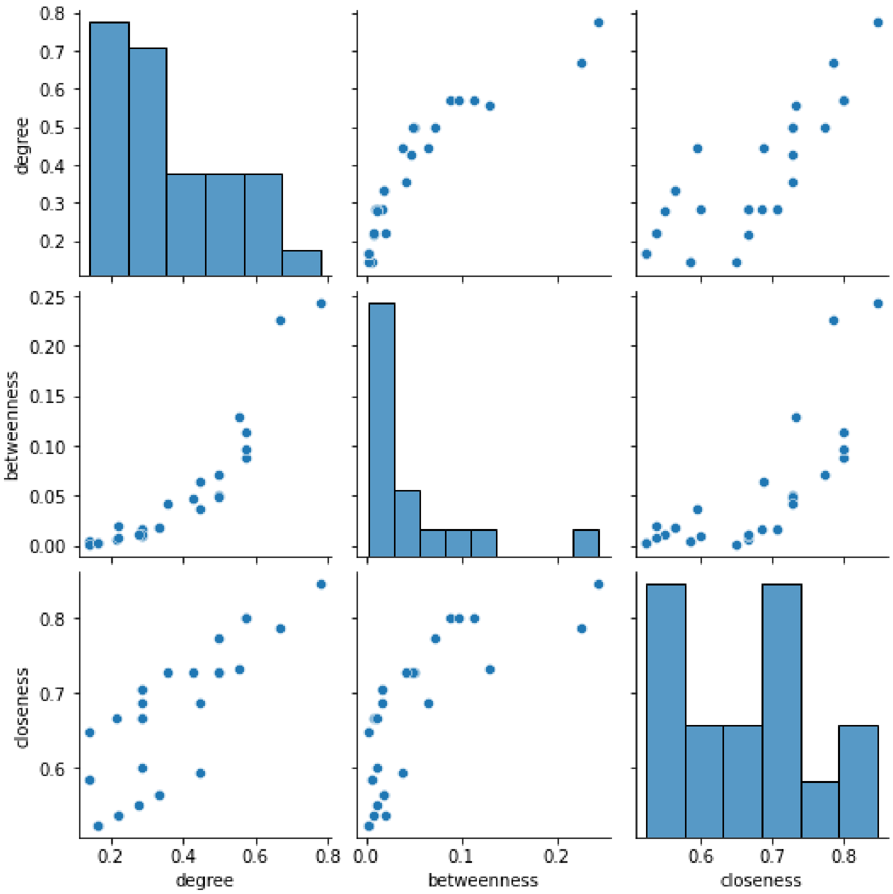

Fig. 4. The result of the centralisation of values, measures and their interrelationship.

tralisation measures for bipartite graphs it is necessary to determine the degree of the vertex which is divided by the maximum possible degree. The main difference between unipartite networks and bipartite networks is that in unipartite networks the maximum degree of a vertex is $n-1$, where n is the total number of vertices in the graph, while in a bipartite graph the maximum degree of a vertex is only the total number of vertices in the opposite set, i.e. the maximum degree for a woman in our graph is the number of events. In order to use functions to compute centrality measures for bipartite graphs in NetworkX, we need to pass a set with all nodes in one bipartite set as an argument. The result is a table with women and measures in the rows and their centrality values, measures, and relationships to each other in the columns. We will display some of the data in the table, and from the full information we will draw a figure with 9 schemes (Table 1).

Table 1. Result of centralisation of values, measures and their interrelation.

Name	degree	betweenness	closeness
Eleanor Nye	0.285714	0.009444	0.666667
Evelyn Jefferson	0.571429	0.096585	0.800000
Ruth De Sand	0.285714	0.016783	0.705882
Verne Sanderson	0.285714	0.015738	0.705882
E14	0.166667	0.002241	0.523810
E11	0.222222	0.019665	0.536585
E12	0.333333	0.018094	0.564103
E10	0.277778	0.011442	0.550000

The graphs (Fig.4) show that people are more interconnected than they think, and can overlap interests and be in the same event without agreeing beforehand. Such data allows us to assess people's behavioural patterns and relate them to events.

Conclusions

As a result of the work done, a bipartite network analysis was carried out, the behavioural model of a group of people was evaluated and the centralisation of values, measures and their interrelationships were calculated on their basis. It was determined that the bipartite model strengthens the process analysis, as there is an opportunity to build a network of works, to identify the most relevant actors, groups of actors.

Based on the results, it emerged that people's behavioural patterns are the same: even if people do not know each other, they may have crossed paths and shared acquaintances. Thanks to NetworkX, it was determined that it was possible to assess a person's behavioural pattern by dividing them into separate groups.

Thus, with data about the person and the event they attended and the product they used, it is possible to create targeted advertising, find new customers and be able to analyse the environment at public events.

Data Availability

Codes and data have been deposited in GitHub (https://github.com/adelllina/Mini-project).

References

1. J Harris, JL Hirst, M Mossinghoff, Combinatorics and graph theory. *Undergrad. texts mathematics*, 381–391 (2008).
2. S Hennemann, I Liefner, Proposing bipartite network analysis for the evaluation of regional innovation systems-/-regions, actors, and content. *Proc. 5th Int. Conf. on Innov. Manag. 2008* **5**, 224–233 (2008).
3. SP Borgatti, MG Everett, Network analysis of 2-mode data. *Soc. Networks* **19**, 243–269 (1998).
4. M Latapy, C Magnien, N Del Vecchio, Basic notions for the analysis of large two-mode networks. *Soc. Networks* **30**, 31–48 (2008).
5. L Xiong, GZ Wang, HC Liu, New community estimation method in bipartite networks based on quality of filtering coefficient. *Sci. Program.* **2019**, 12 (2019).
6. D Della Porta, M Keating, Approaches and methodologies in the social sciences. *Camb. Univ. Press. 2008*, 384 (2008).
7. AAV Dossou-Olory, Cut and pendant vertices and the number of connected induced subgraphs of a graph. *Math. Subj. Classif.*, 25 (2019).
8. D Hu, The spectral analysis of random graph matrices. *Enschede: Ipskamp Print.*, 162 (2018).
9. LA Sanchis, Maximum number of edges in connected graphs with a given domination number. *Discret. Math.* **87**, 65–72 (1991).
10. D Vasques Filho, Structure and dynamics of social bipartite and projected networks. *The Univ. Auckl.* **87**, 65–72 (2018).
11. U Brandes, SP Borgatti, LC Freeman, Maintaining the duality of closeness and betweenness centrality. *Soc. Networks* **44**, 153–159 (2016).
12. CL DuBois, et al., 2003 netdata: A collection of network data. (2013).
13. D Allison, BB Gardner, MR Gardner, Deep south: A social anthropological study of caste and class. *Chicago: Univ. Chic. Press.* (1941).
14. LC Freeman, Finding social groups: A meta-analysis of the southern women data. *Dyn. Soc. Netw. Model. Analysis: Work. Summ. Pap.* (2003).

(1) (2) (3) (4) (5) (6) (7) (8) (9) (10) (11) (12) (13) (14)