Lecture 7: Modes and bipartite networks

Eduardo López

Up to this point, we have considered each network on its own. Data we encounter already encodes the relationships as being pairwise. However, in many circumstances, the origin of the pairwise relationship is because there is yet another underlying (prior) relationship playing a role. Examples of such situations involve affiliations and co-location, where two individuals are connected through their connections to an object of another class. In this lecture, we establish some fundamentals about this topic.

1 Modes

The notion of a mode in this context of network analysis refers to the fact that there can be more than one distinct class of object in a network. Furthermore, modes require that links in a network be established between a node of a given mode with a node of a different mode.

A rather simple example of this is illustrated in the game called *Six degrees of Kevin Bacon*. Here's an extract from my undergraduate textbook:

"Here's a game you can play with the name of any actor/actress i you can think of. The objective is to try to identify a movie A in which both Kevin Bacon and i have been in together. If that fails, you can identify a movie B in which i has acted alongside actor/actress j, and then identify movie C in which j and Kevin Bacon have been in together. If that fails, then you can try to find a longer relation, from i to j to h to Kevin Bacon, via some movies D, E, and F. You can continue this approach until you find a long connection between i and Kevin Bacon, or none at all. This game, purportedly created by Brian Turtle [1], is called the "six degrees of Kevin Bacon". The number of movies you need to identify is equal to the Bacon number. The shorter the path to Kevin Bacon (the smaller the Kevin number), the better you do. An illustration of the game is given in Fig. 1. And if you're

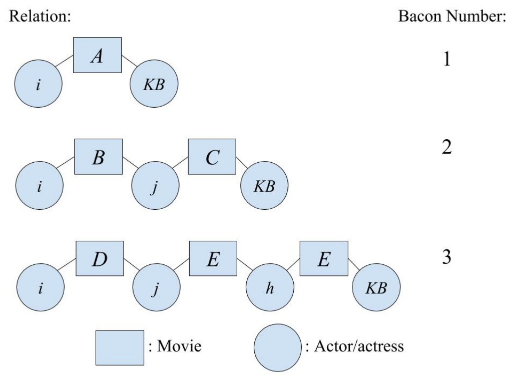

Figure 1: Illustration of the game six degrees of Kevin Bacon.

really curious, go to The Oracle of Bacon (`http://oracleofbacon.org/`*) and type in the name of any actor/actress in the IMDb database to find how they (potentially) connect to Kevin Bacon.*

How is this game relevant to networks and paths in networks, the topic of this chapter? Well, it turns out that the Bacon number is a measure of distance in a network. It tells us how many links need to be crossed to travel from one node of the network to another (in case it is not completely clear, just as for co-authorship, actors are connected into a network by the same rules as authors through a paper).

By the way, the first version of this kind of number is not related to Kevin Bacon, but to a famous Hungarian mathematician named Pál Erdős, who we will talk about again given his important contributions to graph theory. A low Erdős number is a good thing!"

The key point about this description is that actors are related to one another through the movies that they co-act on, not directly. The issue here is the question being asked. We are not concerned with whether or not two actors may be friends, may be related, or live on the same block. We care about how distant the actors are on the basis of the movies they participate in. Thus, movies represent one mode of node, actors another.

With these notions in mind, we can proceed to write

Definition 1. *A **network possesses modes** when there is more than one class of nodes in the network, and links can only be inserted between nodes of different classes.*

Implicit in this definition is the concept of a node class, but it is also operationally defined within this because of the fact that nodes of the same class cannot connect directly.

2 Bipartite networks

The most common form of multi-mode network is called a bipartite network. As the term suggests, a bipartite network is characterized by exactly two node classes. The Kevin Bacon game occurs in a bipartite network. Before we embark in a more formal discussion of this type of network, let us present a few examples that have received some attention in recent years.

2.1 Recommendation systems

The first example we discuss relates to so-called recommendation systems. These are the systems that work *under the hood* in websites such as Amazon or Netflix, which use the user information that connects individuals to the items they select (movies, books, etc.) to learn how to provide meaningful recommendations to other users on the basis of their similarity to others.

Considering the behavior of a user i to be the links this user establishes over time, the link pattern can then be compared to the link pattern of another user j. For instance, user i may have watched movies α, γ, θ, which user j could have watched β, γ, θ. If some score function suggests that i and j are similar enough, the recommendation system is likely to suggest to user i to explore movie β and to user j to explore movie α.

An important component of effective recommendation systems is the deployment of a useful score function that provides a way to find the needle in the digital haystack of information living in very large datasets. In other words, many users may look alike, but choosing the ones that provide the best predictive power to make recommendations to other users is critical for a good recommendation system. In practical terms such algorithms can mean the difference between a successful and a failing business.

2.2 Affiliation networks

This is a large topic in itself and we will hardly do it any justice. Nevertheless, it is important to introduce it as another pathway into bipartite networks.

Think about what societies, organizations, institutions, etc. you belong to. For instance, taking our class is a good example. The relationship that we all have with each other in this class is that of an affiliation. We are all part of the class, and therefore, we have connections. Fundamentally, we do not connect with one another, but rather with the class. After this class ends, it is likely that most of us will disengage.

There are very well known datasets in the study of social networks which are affiliation networks. Perhaps the most classic is the one by Galaskiewicz [2] which focuses on the varied affiliations of a set of CEOs of different companies and clubs, and how these affiliations seem to play a role in decision making even if the CEOs manage different companies.

Another example are people that attend the same event. In a group of people, some may attend various parties, where not all the parties have the same guests.

2.3 Jobs and workplaces

Schmutte [3] uses employers and employees as the two modes of a bipartite network to attempt to develop a structural understanding of the job market. In this example, the author finds that the economy is roughly partitioned in a way that very few transitions occur between certain parts of the labor market.

3 A more complete formulation

In order to formalize some of these concepts for bipartite networks, we introduce some notation and new definitions.

First, let us state the fact that because there are two classes of nodes, we need to define two vertex sets. In addition, since links only occur between the two sets, the edge set is constructed by taking an element of each set. Thus,

Definition 2. *A **bipartite network** H is a collection of two node sets*

$$\begin{aligned} V_e(H) &= \{\alpha, \beta, \dots, \omega\} \\ V_a(H) &= \{i, j, \dots, q\} \end{aligned} \tag{1}$$

and a possibly non-empty edge set

$$E(H) = \{(\gamma, h), \dots\}, \tag{2}$$

where $\gamma \in V_e(H), h \in V_a(H)$.

The names V_e and V_a refer to the vertex set of entities and of actors.

Many of the concepts already defined for a normal network are also valid for bipartite networks. Of great versatility we also find a link indicator $b_{\gamma,h}$

that specifies whether entity γ and actor h are connected or not. Note, however, that the concept of an adjacency matrix changes in its nature. It can still be defined, but its properties change. The network is no longer square in general. Let us present these definitions now.

Definition 3. *The* ***link indicator function*** $b_{\gamma,h}$ ***of a bipartite network*** *of two nodes γ, h each belonging to a different mode, is a function defined by*

$$b_{\gamma,h} = \begin{cases} 1 & \textit{if } \gamma \textit{ and } h \textit{ are connected} \\ 0 & \textit{otherwise.} \end{cases} \tag{3}$$

The collection of link indicators above can be put together in an adjacency matrix $\mathbf{B}$.

Definition 4. *The* ***adjacency matrix*** B *of bipartite network H is the matrix defined by the elements*

$$\mathbf{B}_{\gamma,h} = b_{\gamma,h}. \tag{4}$$

It has dimension $\dim(\mathbf{B}) = |V_e(H)| \times |V_a(H)|$.

As in a typical network, there is a definition of degree, but it is mode-relevant. Thus, there is an entity and an actor degree. Fortunately, the way to think about this continues to be very similar to the case of single-mode networks.

Definition 5. *The* ***actor degree*** *of an actor node in a bipartite network H is given by*

$$k_h^{(a)} = \sum_{\gamma \in V_e(H)} b_{\gamma,h}, \tag{5}$$

which can be interpreted as the sum of the γth row of $\mathbf{B}$*. Similarly, the* ***entity degree*** *of an entity γ in a bipartite network H is given by*

$$k_\gamma^{(e)} = \sum_{h \in V_a(H)} b_{\gamma,h}, \tag{6}$$

equivalent to the sum of the hth column of matrix $\mathbf{B}$*.*

Note that the bipartite nature of these networks does lead to certain differences with simple networks. For instance, bipartite networks do not have triangles in the conventional sense. In other words, it is not possible to go from a node along three distinct links and return to the same node.

Trivially, for a path starting at node i, the closet we can get back to i is to within 1 step:

$$b_{\alpha,i} b_{\alpha,j} b_{\beta,j}$$

can be equal to 1, but at best in case $\beta = \alpha$, the path stays at α and has not yet reached i. The same can be said for a path starting at some entity node α.

The key is that to depart and return to one of the modes, the number of steps needs to be an even number. That is because with every additional link of a path, the node mode changes. Therefore two links returns to the mode the path started with.

Along the lines of the path finding method of taking powers of the adjacency matrix, we can determine the counts of all paths of a given length. But this has to be done with some care given the meaning and associated size issues of $\mathbf{B}$. Thus, to find paths of length 2 that start (and end) in actor nodes, we calculate

Theorem 1. *The* ***number of paths of length*** *2 between two actors i and j in a bipartite network is equal calculated by*

$$\# \textit{ paths between } i \textit{ and } j = \sum_{\gamma} b_{\gamma,i} b_{\gamma,j} = \left(\mathbf{B}^T \mathbf{B}\right)_{i,j}, \tag{7}$$

where $\mathbf{B}^T$ is the transpose of $\mathbf{B}$ and $\mathbf{B}^T\mathbf{B}$ is a matrix of $\dim(\mathbf{B}^T\mathbf{B}) = |V_a(H)| \times |V_a(H)|$. Similarly, the number of paths of length 2 between entity α and entity β are given by

$$\# \textit{ paths between } \alpha \textit{ and } \beta = \sum_{i} b_{\alpha,i} b_{\beta,i} = \left(\mathbf{B}\mathbf{B}^T\right)_{\alpha,\beta}. \tag{8}$$

Many other properties apply to bipartite networks that we will not have time to discuss. However, there is one transformation that stems from this last theorem that we do discuss next.

4 Single mode networks

Bipartite networks can have *projections* into a single mode network by using the results of Theo. 1. Focusing on the projection relevant to actors, the matrix product $\mathbf{B}^T\mathbf{B}$ has dimensions $|V_a(H)| \times |V_a(H)|$ and the values of the elements are the number of paths of length 2 between each of the nodes represented by the location of the element. But this is basically the same as writing an adjacency matrix between two nodes in a simple network, and instead of assigning a 0 of a 1 to the matrix element, we would assign an integer-valued *weight*. Similar statements apply to $\mathbf{B}\mathbf{B}^T$ with the focus being on the entities now. The networks emerging this one are called single-mode networks, emphasizing the fact that they originate from bipartite networks (by the way another name for bipartite networks is two-mode networks).

These single mode networks can also be made unweighted if that was desired. One would merely have to make use of Heaveside step function $\theta(u)$, equal to 1 is $u > 0$, and 0 if $u \leq 0$.

To determine fundamental quantities such as the degrees of nodes in these single mode networks, one indeed has to put the step function to use, as degree does not pay attention to the weight of links in a network (a concept we introduce below does address this, though). Therefore, if one was interested in the degree of actor-node i, we would calculate

$$k_i = \sum_j \theta\left(\sum_\alpha b_{\alpha,i} b_{\alpha,j}\right) = \sum_j \theta\left(\sum_\alpha \mathbf{B}^T_{i,\alpha} \mathbf{B}_{\alpha,j}\right). \tag{9}$$

Analogously, for a single mode network focused on the entities

$$k_\alpha = \sum_\beta \theta\left(\sum_i b_{\alpha,i} b_{\beta,i}\right) = \sum_\beta \theta\left(\sum_i \mathbf{B}_{\alpha,i} \mathbf{B}^T_{i,\beta}\right). \tag{10}$$

The concept of weight leads to another node centric property known as the strength of a node, defined as

Definition 6. *The* ***strength*** s_i ***of node*** i ***in a weighted network*** *is given by*

$$s_i = \sum_j \sum_\alpha b_{\alpha,i} b_{\alpha,j} = \sum_j \sum_\alpha \mathbf{B}^T_{i,\alpha} \mathbf{B}_{\alpha,j}. \tag{11}$$

Similarly,

$$s_\alpha = \sum_\beta \sum_i b_{\alpha,i} b_{\beta,i} = \sum_\beta \sum_i \mathbf{B}_{\alpha,i} \mathbf{B}^T_{i,\beta}. \tag{12}$$

The fact that these expressions simplify greatly with the removal of the step function means that we can readily notice the following theorem

Theorem 2. *The strength of single mode nodes calculated from a bipartite network H with adjacency matrix $\mathbf{B}$ are given by*

$$s_i = \sum_j \sum_\alpha b_{\alpha,i} b_{\alpha,j} = \sum_\alpha \sum_j b_{\alpha,i} b_{\alpha,j} = \sum_\alpha b_{\alpha,i} k^{(e)}_\alpha. \tag{13}$$

Equivalently, the strength of the entity mode satisfies

$$s_\alpha = \sum_\beta \sum_i b_{\alpha,i} b_{\beta,i} = \sum_i \sum_\beta b_{\alpha,i} b_{\beta,i} = \sum_i b_{\alpha,i} k_i^{(a)}. \tag{14}$$

5 Hypergraphs

One additional concept we can introduce is that of a hypergraph. This concept emerges as a kind of alternative to the bipartite formulation. A hypergraph is a generalization of a graph in that it expands the concept of a link as pair of nodes, to that of a hyperlink which can encompass various numbers of nodes.

References

[1] https://en.wikipedia.org/wiki/Six_Degrees_of_Kevin_Bacon

[2] Galaskiewicz J, and Wasserman S (1989) Mimetic Processes Within an Interorganizational Field: An Empirical Test. Administrative Science Quarterly 34(3): 454-479

[3] Schmutte IM (2013) Free to Move? A Network Analystic Approach for Learning the Limits of Job Mobility (Submitted)