GRAPHS, RANDOM GRAPHS, AND AGENTS DECISIONS

BLEEXY TEAM

A graph $G = (V, E)$ consists of a set of vertices V and a set E of edges between vertices, as in Figure 1. Two vertices are called **neighbors** if they are joined by an edge in G. For a given vertex v, the number of neighbors of v is called the degree of v and it is denoted by $d(v)$.

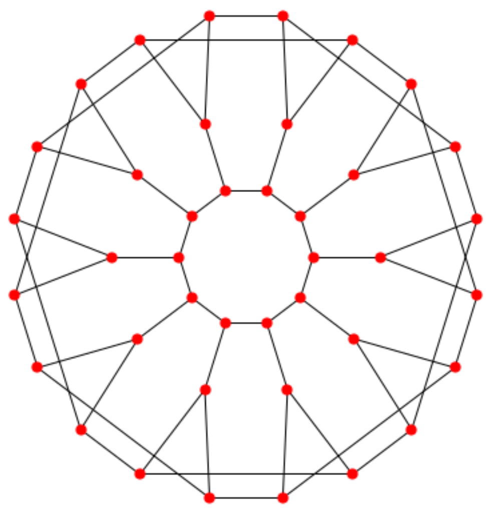

FIGURE 1. An example of a graph

Some everyday example of large graphs (also called networks) are the web-pages that one visits together with the third-party companies that follow the user. The users of any social media company form a graph, where the degree of a vertex representing a user is the number of friends that user has on that particular platform. We can use graphs also to represent shoppers that buy products on a specific website, where two of them are connected if they bought the same product, for example. One can also use graphs for the users who provide reviews for products on a website, where, for example, two users are connected if they provide a review for the same or similar product.

We assume that both V and E are finite set of high cardinality. We let n be the cardinality of V. Then a graph G can be described by a $n \times n$ matrix I, called the incidence matrix, that has entries 0 or 1. If the (i, j) entry of the matrix is 1, then there is an edge from i to j; if the entry is 0, then there is no edge between i and j. We assume that the graph is not directed; that is, if the (i, j) entry of the incidence matrix is 1, then so is the (j, i) entry. Given two vertices v and w in V, we say there is a path between v and w in V if there is a

Date: September 15, 2021.

sequence of vertices $v_1, v_2, \ldots, v_k$ such that $v_1 = v$, $v_k = w$ and there is an edge e_i between v_i and v_{i+1} for all $i = 1, \ldots, n-1$. We write $\mathbf{e} = e_1 e_2 \ldots e_k$ for the sequence of edges and we call it a path from v to w. A graph is called **connected** if there is a path between any two vertices of the graph. A sub-graph of G is a graph G' whose set of vertices is a subset of V and whose set of edges is a subset of E. A connected component of a non-connected graph is a sub-graph that is connected.

For example, a connected component of a social media platform consist of all the users that are friends with the friends of ones friends. A connected component of Netflix users could be all the users that watched all seasons of "Money Heist".

Of more interest are probably graphs that describe how many times two users or agents make the same or similar decision. For example, how many times two Facebook users like the same news stories, how many time two users watched the same movie or TV show on Netflix, or how many times two users provided the same rating for various products. These connections can be described by probabilities and we are led to consider random graphs.

Random graphs were defined in late 50's by Erdős and Rényi in [2] and they have been used recently to study large networks (see, for example, [1]). One of the original random graph models $G(n, p)$ in [2] was defined as a graph G with n vertexes such that there is an edge between two vertexes with probability p. Based on real life examples, one is led to consider a generalization of this model: a random graph $G(n, IP)$ consists of a set of n vertices V where we replace the incidence matrix with a matrix of probabilities IP whose entries (i, j) contain the probability p_{ij} that there is an edge from the vertex i to the vertex j. Given a matrix of probabilities IP, graphs generated from this matrix will differ in the number of edges and which vertices are connected as exemplified in Figure 2.

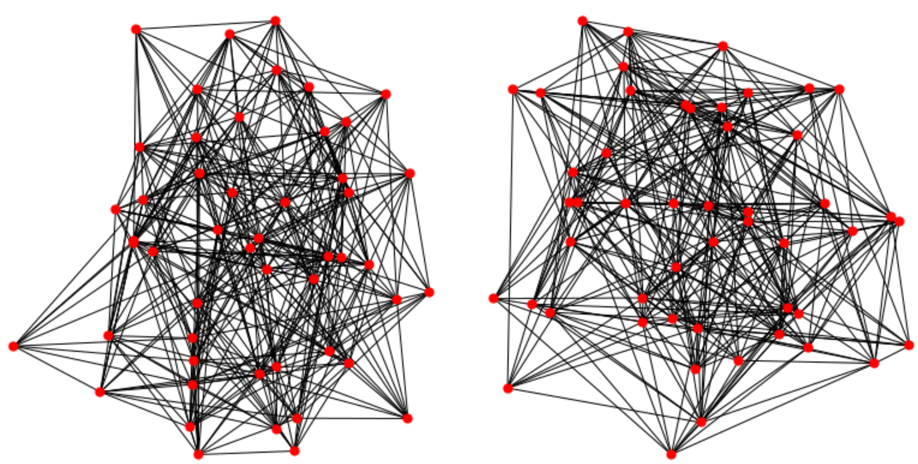

FIGURE 2. Two graphs generated from the same probability matrix

The probabilities p_{ij} can be interpreted as the percentage which two agents/users made the same choice or rating. The closer to 1 it is, the more likely is that the two agents will make the same decision in the future as well. If a company wants to push new products or shows to users, that it wants to look for users that are interconnected via high probabilities.

However, if users get paid for their decisions or choices and the payment depends on how close the user is to the average decision, then the company must monitor that a group of users does not monopolize the decisions and choices for their own benefit. In this case, choosing agents connected with low probabilities is better. In addition, the company might want to study how connected are different group of users. There are a few characteristics of graphs and random graphs that can provide some clarity when trying to determine how users group in a network. We describe next a couple of them and we will discuss more in a future whitepaper.

The clustering coefficient is a measure of the degree to which vertices in a graph tend to cluster together. Mathematically, it is the quotient

$$C = \frac{3 \cdot \text{the number of triangles in the graph}}{\text{number of connected triplets of vertices}}.$$

Three vertices $\{i, j, k\}$ in a graph are connected if there is an edge from i to j and an edge from j to k. The vertices form a triangle if, in addition, there is an edge from k to i as well. For example, if two of your friends are also friends together, then the three of you would form a triangle. If C is close to 1 then most of the vertices in a graph are part of at least one triangle.

The connectivity threshold of a random graph or subgraph is the matrix IP such that if at least one of the values of the probabilities p_{ij} increases the graph becomes connected. That is, if at least two users/agents make one more of the same decision, either watch the same movie, buy the same product, or provide the same rating for a product, then the graph of these users becomes connected. Some companies would like for the graph to become connected, while other companies will try to avoid reaching the connectivity threshold in order to avoid users taking over their networks.

One component of the Bleexy ecosystem that uses the theory of random graphs is the validation process. Central to Bleexy's concept of helping business with fair and useful reviews for their products is the procedure in which a group of agents, called "validators", selected from a pool of agents based on their skills and interests, review and provide scores to new pieces of content attached to various products. Pieces of content can be photos, short video clips or written reviews, for example. After the selected group of validators review the piece of content assigned to them, each agent provides a score. Based on the average of these scores, the piece of content can be *validated*, that is, accepted into the Bleexy network, or not. Validators whose scores are close to the final validation score receive a boost in their trustworthiness scores, while those whose scores are way off from the validation score, will see their trustworthiness scores diminish. Having a higher trustworthiness score leads to a higher chance of being selected for a validation process in the future and, hence, to higher earnings. This leads to quality and trusted content for each product, and, in turn, to higher sales for producers and sellers. Having a highly trusted network of content creators and validators is a key advantage of Bleexy. Moreover, the validation algorithm has built in mechanisms to check against hostile attacks. To prevent validators to team up in order to increase their trustworthiness scores and to prevent a takeover of the Bleexy network by a hostile group, the algorithm uses "randomness" in the validators selection process. Validators are selected based in part on the likelihood that they will provide a similar score or not. Specifically, validators are grouped in a random graph based on the percentage computed from how many times they provided similar scores for different pieces of content. Bleexy uses these random graphs to randomly select validators in order to guarantee that no agent can control which validators are selected at any given moment in time. By regularly computing characteristics of the random graphs such as the clustering coefficient or the connectivity threshold described above, Bleexy can detect and prevent the grouping of validators or hostile takeover of the network. The clustering coefficient allows Bleexy to monitor whether validators are attempting to form cliques. The connectivity threshold combined with other characteristics of the random graph allows Bleexy to detect and prevent a hostile takeover and teaming up of actors and make sure that only quality and trusted content is validated.

References

[1] Shankar Bhamidi, Suman Chakraborty, Skyler Cranmer, and Bruce Desmarais. Weighted Exponential Random Graph Models: Scope and Large Network Limits. *J. Stat. Phys.*, 173(3-4):704–735, 2018.

[2] P. Erdős and A. Rényi. On random graphs. I. *Publ. Math. Debrecen*, 6:290–297, 1959.