Affiliation Networks

SILVIO LATTANZI, D. SIVAKUMAR

Presented By:
Aditi Bhatnagar

Under the guidance of:
Augustin Chaintreau

Objective

- To provide a simple, realistic and *Mathematically* tractable model which are *algorithmically* useful and:-

a) explains all properties of social networks
b) explains *densification* and *shrinking diameter.*

- Modular approach to connecting random graph paradigms to structural consequences.

Motivation

- Internet and Web Graphs

1. Faloutsos:- The degree distribution of the Internet graph is heavy-tailed, and roughly obeys a "power law".
2. Barabasi And Albert :- A network evolves by new nodes attaching themselves to existing nodes with probability proportional to the degrees of those nodes.
3. Broder:- Besides power-law degree distribution, the web graph consisted of numerous dense bipartite subgraphs.

- Preferential attachment and edge copying are two basic paradigms that both lead to heavy-tailed degree distributions and small diameter.

Few Basic Definitions

- **Preferential attachment:-** "Rich getting richer", Yule Process
- **Edge Copying:-** each new vertex picks an existing vertex as its "prototype," and copies (according to some probabilistic model) its edges.
- **Bipartite Graph:-** graph whose vertices can be divided into two disjoint sets U and V such that every edge connects a vertex in U to one in V; that is, U and V are independent sets.
- **Duality of Actors and Groups/Societies**:- Groups can be described as collections of actors affiliated with it and actors can be described as collections of groups with which they are affiliated.
- **Heavy-Tailed Distribution**:- They have heavier tails than the exponential distribution.

Motivation

- Small World Graphs

1. **Kleinberg :-** Nice Starting point to analyze social networks. Transitive Friendship is observed.

2. **Limitations:-**

a) Not applicable in developing an understanding of real social networks.

b) Static Model.

Densification and Shrinking Diameter

1. <u>Leskovec</u>

a) Real-world networks became denser over time and their diameters effectively decreased over time.

b) Community guided attachment and Forest fire model.

2. Limitations:-

a) Complex model.

b) Do not powerfully analyze degree distribution, densification, and shrinking diameter simultaneously.

Affiliation Networks

Social networks there are two types of entities —***actors*** and ***societies*** — that are related by affiliation of the actors in the societies.

1. *Affiliation networks* are the social network among the actors that results from the bipartite graph.
2. Obtained by “folding” the graph- replacing paths of length two in the bipartite graph among actors by an (undirected) edge.
3. One or more common or shared affiliations.

Affiliation Networks

- *Affiliation network B* and its *folding G* on *n* vertices is produced, the resulting graphs satisfy the following properties:-

1. B has a power-law distribution, and G has a heavy- tailed degree distribution.
2. Wit some mild condition on ratio of the expected degree of actor nodes and society nodes in B, the graph G has superlinear (function that grows faster than a linear one) number of edges.
3. The effective diameter of G stabilizes to a constant.

Algorithmic Benefits and an Application

Main Question: What they want to achieve with this model?

Answer:

In a large random set R:-

a) paths from arbitrary nodes to nodes in R then we can sparsify the graph.
b) preserve all shortest distances to vertices in R.
c) Sparsify the graph to have a linear number of edges.
d) stretching distances by no more than a factor given by the ratio of the expected degree of actor and society nodes in the affiliation network.

- Example :- The affiliation network is the bipartite graph of queries and web pages (urls), with edges between queries and urls that users clicked on for the query.

Model

B(Q,E)

- Fix two integers $c_q, c_u > 0$, and let $\beta \in (0, 1)$.
- At time 0, the bipartite graph $B_0(Q,U)$ is a simple graph with at least $c_q c_u$ edges, where each node in Q has at least c_q edges and each node in U has at least c_u edges.
- At time $t > 0$:
- (**Evolution of Q**) With probability β:
- (Arrival) A new node q is added to Q.
- (**Preferentially chosen Prototype**) A node $q' \in Q$ is chosen as prototype for the new node, with probability proportional to its degree.
- (**Edge copying**) c_q edges are "copied" from q'; that is, c_q neighbors of q', denoted by $u_1, \ldots, u_{c_q}$, are chosen uniformly at random (without replacement), and the edges $(q, u_1), \cdots, (q, u_{c_q})$ are added to the graph.
- (**Evolution of U**) With probability $1 - \beta$, a new node u is added to U following a symmetrical process, adding c_u edges to u.

MODEL

G(Q,E)

- Fix integers cq, cu, s > 0, and let $\beta \in (0, 1)$.
- At time 0, G0(Q,E) consists of the subset Q of the vertices of B0(Q,U), and two vertices have an edge between them for every neighbor in U that they have in common in B0(Q,U).
- At time t > 0:
- (**Evolution of Q**) With probability β:
- (**Arrival**) A new node q is added to Q.
- (**Edges via Prototype**) An edge between q and another node in Q is added for every neighbor that they have in common in B(Q,U) (note that this is done after the edges for q are determined in B).
- (**Edges via evolution of U**)
- With probability $1 - \beta$:
- A new edge is added between two nodes q1 and q2 if the new node added to $u \in U$ is a neighbor of both q1 and q2 in B(Q,U).
- (**Preferentially Chosen Edges**) A set of s nodes qi1 , . . . , qis is chosen, each node independently of the others (with replacement), by choosing vertices with probability proportional to their degrees, and the edges (q, qi1), . . . , (q, qis) are added to G(Q,E).

EVOLUTIONOF THE DEGREE DISTRIBUTIONOF AFFILIATION NETWORKS

- Degree Distribution of B(Q,U)
- ***Result 1-*** The degree distribution of vertices in both Q and U satisfy power laws.
- ***Result 2*** - most of the edges of B added "later" in the process have their end points pointing to a low-degree node.

- ***Theorem-*** For the bipartite graph B(Q,U) generated after n steps, almost surely, when $n \to \infty$, the degree sequence of nodes in Q (resp. U) follows a power law distribution with exponent $\alpha = -2 - cq\beta/(cu(1-\beta))$

PROPERTIES OF THE DEGREE DISTRIBUTIONS of G (Q,E)

- The degree distributions of the graphs G(Q,E) is heavy-tailed.
- Most the nodes have degrees in $\Theta(1)$.

DENSIFICATION OF EDGES

- The number of edges in the graph G(Q,E) is $\omega(|Q|)$.
- ***Theorem*** :- If $c_u < \beta/1-\beta\ c_q$ the number of edges in G(Q,E) is $\omega(n)$.
- Proof …

SHRINKING/STABILIZING OF THE EFFECTIVE DIAMETER

- Effective Diameter:-For 0 < q <1, we define the q effective diameter as the minimum de such that, for at least a q fraction of the reachable node pairs, the shortest path between the pairs is at most de.
- If a person *q* is not interested in any popular topic, and so is not linked to any popular topic in B(Q,U), with high probability at least a friend of q is interested in a popular topic.
- ***Theorem***:- If cu < β/1−β cq, the q-effective diameter of the graph G(Q,E) shrinks or stabilizes after time φn, for any constants 0 < φ, q < 1.

SPARSIFICATION OF G(Q,E)

- ***Sparsification Algorithm***

Input: G(Q,E) and a set R of relevant nodes.

1. Initially, label all edges deletable.
2. For each node a ∈ R:

(a) Compute the breadth first search tree starting from node a and exploring the children of a node in increasing order of insertion.

(b) Label all edges in the breadth first search tree of node a as undeletable.

3. Delete all edges labeled as deletable.

Sparsification with a stretching of the distances

- Some bounded stretching of the shortest distance between two nodes is permitted.
- Theorem:- There is a polynomial algorithm that, for any fixed cu, cq, β, finds a graph G′(Q,E′) with a linear number of edges, where the distance between two nodes is at most k times larger that the distance in G(Q,E) and in ˆ G(Q, ˆE), where k is a function of cu, cq, β.

FLEXIBILITY OF THE MODEL

- Instead of generating only one bipartite graph B(Q,U), a list $B_0(Q,U), \cdots, B_k(Q,U)$ of bipartite graphs 6 are generated. At the same time the multigraph G(Q,E) evolves in parallel; besides "folding" length-2 paths in $B_0, \cdots, B_k$ into edges, we also add to G(Q,E) a few preferentially attached neighbors.
- Instead of "folding" length-2 paths in B into edges, for every pair of nodes in Q and every shared common neighbor $u \in U$ between them, we randomly and independently place an edge between the nodes in G (Q,E) with probability proportional to the reciprocal of $d(u)^\alpha$, where $d(\cdot)$ denotes degree $0 < \alpha < 1$.

Thank You