$\alpha$-bins and Random Threshold Graphs

Sudipta Saha
Indian Institute of Technology
Kharagpur – 721302, India
Plan of the discussion

- Alpha-bipartite networks
  - Introduction

- Edges
  - Creation technique and meaning

- Our aim
  - Special structure - One-mode projection

- Modeling real life constraints
  - Thresholds

- Works already done

- Further research issues
  - Understanding the structure

- New approach
  - Random threshold graphs
  - Graphs as binary strings

- Evolution of random threshold graphs

- Present status of the work

- Discussion
Alpha-bipartite networks (α-bins)

- Many real life systems comprise of two distinct entity sets
  - Bipartite Network
    - Two partitions
  - A special property
    - Fixed partition
    - Growing partition
  - Alpha-Bipartite Network (a name originated from linguistics)
  - Examples
    - Word - letter
    - Gene - codon
    - User - group
    - Railway - station
    - Protein – amino-acid
    - People - place
    - Event - person
    - Movie - actor
    - Paper - author
Alpha-bipartite networks (α-bins)

Set of English words/ sentences

Set of 64 codons

Gene-codon

Set of genes

Set of Proteins

Set of users/ people

English alphabet

Word-letter

People-place / User-group

Groups / Places of interest

Protein-amino-acid

Set of 20 amino acids
Interpretation of edges in $\alpha$-bins

- Member of a growing set may connect to one or many nodes in the fixed set

- Meaning / implication of multiple connections
  - Subject / topic similarity
    - Groups in user-group network (online social network)
    - Places in people-place network
    - Authors in paper-author network
  - Co-occurrence
    - Codons in gene-codon network
    - Amino-acids in protein-amino-acid network

- Information flow
  - Groups in user-group network in online social networks by posting / sharing of information
  - Among places in people-place network

- Flow of disease / opinion
Alpha-bipartite networks (\(\alpha\)-bins)

- Selection of an element of the fixed set by an element of the growing set
  - Fully Preferential
    - The probability that a node is selected at time step \(t\) is proportional to the number of connections it has already acquired before \(t\)

\[
\Pr\{d_{t+1}(i) = d_t(i) + 1\} = \frac{d_t(i)}{\sum_{j=1}^{N} d_t(j)}
\]

- \(d_t(i)\) = degree of the \(i^{th}\) element in the fixed set after \(t\) number of nodes have joined the growing set
- \(N\) = cardinality of the fixed set
Alpha-bipartite networks (α-bins)

- Popularity / attractiveness
  - A random variable - denoted by $\theta_i$
  - Asymptotic growth rate of the degree of a node in the bipartite network
  - Can be sampled from Beta distribution / Dirichlet distribution

![Diagram showing probability distribution and network partitions]
Alpha-bipartite networks (α-bins)

- Selection of an element of the fixed set by an element of the growing set
  - Fully Random selection

\[ \Pr\{d_{t+1}(i) = d_t(i) + 1\} = \frac{1}{N} \]

- \(d_t(i)\) = degree of the \(i\)th element in the fixed set after \(t\) number of nodes have joined the growing set
- \(N\) = cardinality of the fixed set

- All the nodes acquired almost the same number of connection

- Fixed partition
- Growing partition
Alpha-bipartite networks (α-bins)

- Selection of an element of the fixed set by an element of the growing set
  - Fully preferential and fully Random selection can be combined in a single formula

\[
\Pr\{d_{t+1}(i) = d_t(i) + 1\} = \frac{d_t(i) + \delta}{\sum_{j=1}^{N} (d_t(i) + \delta)}
\]

- \(d_t(i)\) = degree of the \(i^{th}\) element in the fixed set after \(t\) number of nodes have joined the growing set
- \(N\) = cardinality of the fixed set

\(\delta\) is a regulating parameter
  - \(\delta = 0\), Fully preferential case
  - \(\delta = \text{Very large number}\), Fully random case
**One-mode projection of α-bins**

- **Focus** - understanding the connectivity / relationships among the nodes in the fixed set – indirectly the growing set
  - Special structure – one-mode projection
    - An induced uni-partite graph on the fixed partition
  - Formation – one possible way
    - An edge: both nodes are connected to some common element of growing set
    - Weight of an edge: count of the distinct paths of length 2 via growing partition

**Diagram**

- Set of people
- Set of places
- Joining of a node in the growing set
- A clique in the one-mode projection
Construction of one-mode projection

Growing set V

Fixed set U

Projection – multiple edges
Construction of one-mode projection

Growing set $V$  
Fixed set $U$  
Projection – multiple edges
Construction of one-mode projection

Growing set V

Fixed set U

Projection – multiple edges
Construction of one-mode projection

Growing set V

Fixed set U

Projection – multiple edges
Construction of one-mode projection

Fixed set $U$

Growing set $V$

Weighted projection

Degree of a node = sum of weights of the edges
Construction of one-mode projection

- Different ways to form one-mode projection (induce a graph on one partition)
  - Multiple connections to a single node in the fixed set can be assumed to be a single connection

  Examples
  - An user in the online social network may post different messages in a single group multiple times
  - A single protein may be expressed by a certain gene multiple times

But in one-mode projection formation we may ignore and consider only single connection
Construction of one-mode projection

Group-activity by users

Projection – ignoring multiple connections

Projection – considering multiple connections
Various ways to form one-mode projection

- Different other ways to form one-mode projection
  - Probabilistic formation of the edges

Examples:
- People-party bipartite network
  - Coming to a specific event by two different persons
  - May not lead to relationship formation between them
- Similarly – movie - actor bipartite network

In one mode projection we can reflect this by using a probability distribution to connect any two persons coming to the same party

Visiting / connecting two distinct places may not deterministically lead to a connection formation between the two nodes in the fixed set.
Various ways to form one-mode projection

- Projection - deterministic
- Projection - probabilistic
One-mode projection after large growth

- Understanding the connectivity / relationships among the nodes in the fixed set after a large growth of the growing set
  - Asymptotic status depends on the initial condition
  - Attachment policy etc
    - It usually becomes a complete weighted graph

- Previous studies in this direction yielded
  - Asymptotic degree distribution *
  - Effect of randomness in the attachment policy *
  - Many other interesting properties and observations

Real life constraints
- Generic constraints - **Time** and **Energy** - with respect to message dissemination in various distributed systems
  - Mobile networks
  - User groups
- Specific – buffer time
  - Buffer-equipped DTN

A simple scenario of distributed communication among different places

From the perspective of dissemination of message under constrained resource, the probability that a message will pass from Place A to B by some person, is higher than that between both A and C or B and C.
Real life constraints can be modeled by **thresholds** on the one-mode projection.

**Thresholded one-mode projection**
- A threshold value
- Prune those edges from the one-mode projection (complete graph) whose edge weights are below the threshold value
- Un-weighted graph
Works done on thresholded one-mode projection

- Previous and ongoing works in this direction
    - Threshold represents a specific constraint \`Message buffering time\’ in places for enhancing indirect communication in DTN
    - Yields the notion of optimal buffer time
    - Detailed analysis of the thresholded one mode projection
    - Understanding the communication / information flow between user-groups
    - Comparing with real data – youtube, flickr and livejournal
Research issues

- Detailed analysis of the thresholding process
  - For threshold, Δ=0
    - Complete Graph
  - For sufficiently large value of threshold
    - All nodes are isolated
  - Graph configurations for other threshold values
    - Have not been investigated yet
- Significance
  - It is essential to understand structure of the graph at various threshold value

A new approach: Thresholded one-mode projections as random threshold graphs
Threshold graphs and random threshold graphs

- **Threshold Graphs** (Chvátal and Hammer, 1977)
  - Nodes
    - Associated with a real number
    - Example—popularity of an actor, richness (wealth) of a person
  - Edge between two nodes
    - Sum of the node weights crosses a certain threshold $\Delta$
      - $W_i + W_j > \Delta$
    - “Rich persons always know each other”

- **Random Threshold Graphs** (Diaconis et al, 2009)
  - Weights are chosen as independent and identically distributed random variables (i.i.d.) from some probability distribution

- **Example**
  - Nodes with weights 310, 240, and 100
  - Edges with weight $\Delta = 300$
Thresholded-one-mode projection as threshold Graphs

- Previous analysis on one-mode projection yielded
  - The edge weight between two nodes to be $c\theta_i \theta_j$
    - Where $\theta_i$ and $\theta_j$ are two real numbers associated with the two nodes $i$ and $j$ respectively (Popularity / attractiveness)
    - $\theta_i$'s are random variables following beta distribution
    - $c$ is a constant
  - Thresholded one-mode projection
    - An edge exists between node $i$ and $j$ if $c \theta_i \theta_j > \Delta$

- Generalized threshold function
  - Any monotonically increasing and symmetric function can be used to define a threshold graph
  - Product can be easily transformed to addition by taking log

- Any thresholded one mode projection can be thought of as a Random Threshold Graph associated with a specific threshold value
  - All existing theories of random threshold graphs are applicable
Threshold graphs and random threshold graphs

- Applicability - detailed study of the alpha-bins
  - Under various constraints
  - Largest connected component
  - Structural information

- Example:
  - Information flow
    - Delay Tolerant Network
    - User-group network
  - Flow of opinion
  - Protein/Gene
    - Stickiness
    - co-expression

Constraint is modeled by some value of threshold
Which configuration is more likely?
Any threshold graph can be built by sequentially inserting nodes of two types

- Either isolated
  - Denoted by 0
- Dominating
  - Denoted by 1

One-to-one correspondence between a Random Threshold Graph with N nodes and a binary string of length N-1

- The first vertex can be denoted by either 0 or 1
- Easier to deal with strings than graphs

A hierarchical structure

Example: A possible threshold graph with 6 nodes represented by 6 or 5 bit binary string

\[ \begin{array}{cccccc}
0 & 1 & 1 & 0 & 1 & 1 \\
\end{array} \]

\[ N = 5 \]

\[ \begin{array}{cccccc}
0 & 1 & 1 & 0 & 1 & 1 \\
\end{array} \]

\[ \text{An isolated vertex} \]

\[ \text{A dominating vertex} \]

\[ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ \text{and} \ 1 \ 1 \ 1 \ 0 \ 1 \ 1 \]

\[ \text{represent the same threshold graph} \]
Construction of threshold graph

<table>
<thead>
<tr>
<th>A threshold graph</th>
<th>Equivalent binary string</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="A" alt="A" /></td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

The first node can be denoted by 0 or 1
## Construction of threshold graph

<table>
<thead>
<tr>
<th>A threshold graph</th>
<th>Equivalent binary string</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Node A" /></td>
<td>0 0</td>
</tr>
<tr>
<td><img src="image" alt="Node B" /></td>
<td>1 0</td>
</tr>
<tr>
<td>An isolated node (0) is added</td>
<td></td>
</tr>
</tbody>
</table>
Construction of threshold graph

A threshold graph

A dominating node (1) is added

Equivalent binary string

0 0 1

1 0 1

A dominating node (1) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0

1 0 1 0

An isolated node (0) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1

1 0 1 0 1

A dominating node (1) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0

1 0 1 0 1 0

An isolated node (0) is added
Construction of threshold graph

<table>
<thead>
<tr>
<th>A threshold graph</th>
<th>Equivalent binary string</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="https://via.placeholder.com/150" alt="Graph Diagram" /></td>
<td>0010100</td>
</tr>
<tr>
<td></td>
<td>1010100</td>
</tr>
</tbody>
</table>

An isolated node (0) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0 0 0

1 0 1 0 1 0 0 0

An isolated node (0) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0 0 0 1

1 0 1 0 1 0 0 0 1

A dominating node (1) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0 0 0 1 0

1 0 1 0 1 0 0 0 1 0

An isolated node (0) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0 0 0 1 0 0

1 0 1 0 1 0 0 0 1 0 0

An isolated node (0) is added
Construction of threshold graph

A threshold graph

Equivalent binary string

0 0 1 0 1 0 0 0 1 0 0 0

1 0 1 0 1 0 0 0 1 0 0 0

An isolated node (0) is added
Construction of threshold graph

A 12 node threshold graph is represented by

There can be total $2^{N-1}$ distinct threshold graphs from $N$ number of nodes.
Construction of threshold graph

The status at $\Delta = 0$ for $N = 5$ i.e., a 5 node complete graph can be represented by 1111.

The status at very high $\Delta$ i.e., fully isolated status can be represented by 0000.

To understand the configuration/binary string for a given threshold value, we need to identify the full ascending order of the edge weights.

Edges are deleted following the ascending order of the edge weights.
Deletion of edge in threshold graph

- Thresholding implies deletion of edges
  - Degree of a node is related to its ‘attractiveness’
  - The degree of ‘0’ node increases from right to left
  - The degree of ‘1’ node increases from left to right

Higher degree denotes higher attractiveness value
(Higher node weight)

- 0 ... 0  No edge possible
- 1 ... 0  No edge possible
- 1 ... 1  Edge with higher weight
- 0 ... 1  Edges with least possible weights

It can be shown that the ‘01’ sequences represent the least weight edges

Possible edges for deletion
Deletion of edge in threshold graph

Let us take the second occurrence

Interchange the order of insertion of the dominating vertex and the isolated vertex

Each of the vertices will lose one degree which implies deletion of an edge between them
Deletion of edge in threshold graph

- For fixed values of the node weights – we can easily determine this.
- But this is the most difficult job if the node weights are i.i.d. random variables from some distribution.

Consider $N = 4$, nodes with weights (attractiveness) - $\theta_1$, $\theta_2$, $\theta_3$ and $\theta_4$ which are i.i.d. random variables.

- Consider the order ($\theta_1 < \theta_2 < \theta_3 < \theta_4$).
- Relationship between $\theta_1 \times \theta_4$ and $\theta_2 \times \theta_3$ is not easily known.

For 4 variables finding the probability:
- $P\{\theta_1 \times \theta_4 > \theta_2 \times \theta_3\}$ is solvable;
- but becomes very difficult for large $N$. 

But, there can be many occurrences of ‘01’, Which one represents the least weight edge?
Evolution of threshold graphs

- Representing by - Evolution graphs
  - Arranging all possible configurations in course of thresholding of a one-mode projection in the form of a graph where
    - A node: A possible graph configuration
    - Edge: One configuration is possible to achieve from the other configuration by adding or deleting some edge

- Starting from a complete graph with N nodes needs \( \binom{N}{2} \) levels of thresholding
  - Height of the graph is \( \binom{N}{2} \)
  - Number of nodes is \( 2^{N-1} \)
  - But what is the width of a graph?
Evolution graph

\[ \theta_4 > \theta_3 > \theta_2 > \theta_1 \]

- The starting configuration 1111
- The left most bit does not matter – make it 0 always
- Thus we proceed through the pruning the least weight edge
- Whenever there are more than one such possibility, we generate all possible configurations
- Example for a 4 node graph
Evolution graph

$\theta_4 > \theta_3 > \theta_2 > \theta_1$

$\theta_2 \times \theta_3 > \theta_1 \times \theta_4$

$\theta_1 \times \theta_4 > \theta_2 \times \theta_3$

$\theta_4 > \theta_3 > \theta_2 > \theta_1$

N=4

Fully connected

Fully isolated

N=4
N=6
Height is $\binom{6}{2} = 15$
Number of nodes = 32

Evolution graph
Width of the evolution graph

- We have N nodes in the projection
- Each node (ith) can be either dominating or isolated (0 or 1)
- If dominating then
  - Number of edges created at the time of its insertion is same as its position – 1 i.e., (i - 1)
- If isolated then
  - Brings no edge (0)
- For N nodes we have the following Generating function
- This gives us \(^{N}C_2\) coefficients each of which corresponds to the number of possible graphs for a given number of edges (i.e. at a certain level of the evolution graph)

The problem is similar to finding how many ways we can partition an integer N by the integers from 1 to N-1

For N=4, the coefficients are 1 1 1 2 1 1 1
Coefficients (log-linear plots)

Value of the Threshold (Degree of specific constraint)

Number of possible graph configurations / Width of the evolution graph
Coefficients (rescaled linear plots)

Rescaled value of the Threshold (Degree of specific constraint)

Rescaled number of possible graph configurations /
Width of the evolution graph
Problems yet to be solved

- All graph configurations at a certain level do not have equal probability
- Thus all paths also do not have equal probability
- Finding out the typical paths in the evolution graphs
  - Finding the full order of the pair-wise products of N random variables – for few possible distributions of their weights
    - Beta, power-law
Thank You

Sudipta Saha
sudipta.saha@cse.iitkgp.ernet.in
http://cse.iitkgp.ac.in/~sudiptas