The network properties of episodic graphs

Yuwen Zhuang, Vishnu Sreekumar, Mikhail Belkin and Simon Dennis
The Ohio State University

Episodic Structure

- Episodic structure can be studied by using graphs to represent an episodic network.
- Analysis of this structure can be used in future models of episodic memory and context reinstatement.

Basic concepts of graph theory and small-world structure

Small-world Structure:
A small world network possesses the following properties:
i) The small-world effect means most pairs of vertices are connected by a short path through the network.
ii) High "clustering" or "transitivity" means that if two vertices have another neighboring vertex in common, there is a high probability that these two vertices will be connected directly to each other.

Clustering Coefficient:
Global clustering coefficient defined as follows:

$$C=\frac{3\times(\textit{number of triangles on a graph})}{(\textit{number of completed connected triples of vertices})}$$

- A triangle means three vertices are pairwise connected to both of the others.

- A connected triple means a vertex that is connected to an (unordered) pair of other vertices.

- Each triangle contributes to three separate connected triples.

IMDb Graphs

- The IMDb corpus contains 12738 documents.
- The articles focus on celebrity gossip and are 80-400 words long.
- Latent Semantic Analysis is used to construct paragraph vectors. Before that we used Named Entity Tagger to tag plain text with named entities (people / organizations / locations /miscellaneous).
- We construct a matrix whose elements, M_{ij} are given by

$$M_{ij}=\ln(m_{ij}+1)S_j$$

where m_{ij} is the number of times that j_{th} word is found in i_{th} document, and

$$S_j=1+\frac{\sum_{i=1}^{N}P_{ij}\ln(P_{ij})}{\ln(N)}$$

is the weight given to each world, which depends on the information entropy of the word across documents. In the above expression

$$P_{ij}=\frac{m_{ij}}{\sum_{i=1}^{N}m_{ij}}$$

is the probability of the j_{th} word in the i_{th} document, and N is the total number of documents.

Similarity Measure The measure of similarity between documents used is the cosine of the angle between the two vectors.

Image Graphs

Microsoft Research SenseCam

- Microsoft Research SenseCams are used to capture images for several subjects.
- Each subject carried a SenseCam to record the images for their daily lives.
- The SenseCam can take photographs at regular intervals (8 seconds). The sensors in the SenseCam let the camera automatically take pictures when changes in color, light-intensity and temperature detected.

Representation in HSV Space

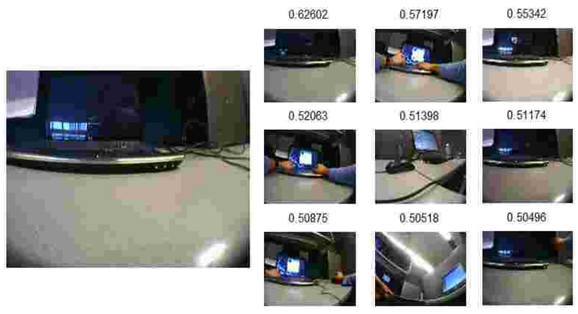
- HSV color space used.
- The hue circle consists of the primaries red, green and blue separated by 120 degrees.
- Quantization of hue requires the most attention. Thus, H is quantized to 30 levels and S & V is quantized to 10 and 3 levels. The quantized HSV space has 900 histogram bins.

Color Histogram and Distance Measure

- Color histogram used.
- Use histogram intersection distance to measure the similarity among images.
- The intersection of histograms h and g is given by:

$$d(h,g)=\frac{\Sigma_A\Sigma_B\Sigma_C\, min(h(a,b,c),g(a,b,c))}{min(|h|,|g|)}$$

where |h| and |g| gives the magnitude of each histogram,which is equal to the number of samples.



Query image and retrieved images using the color histogram intersection method. The distances from the query image are on top of each image.

Graph Construction

IMDb and Image Graphs

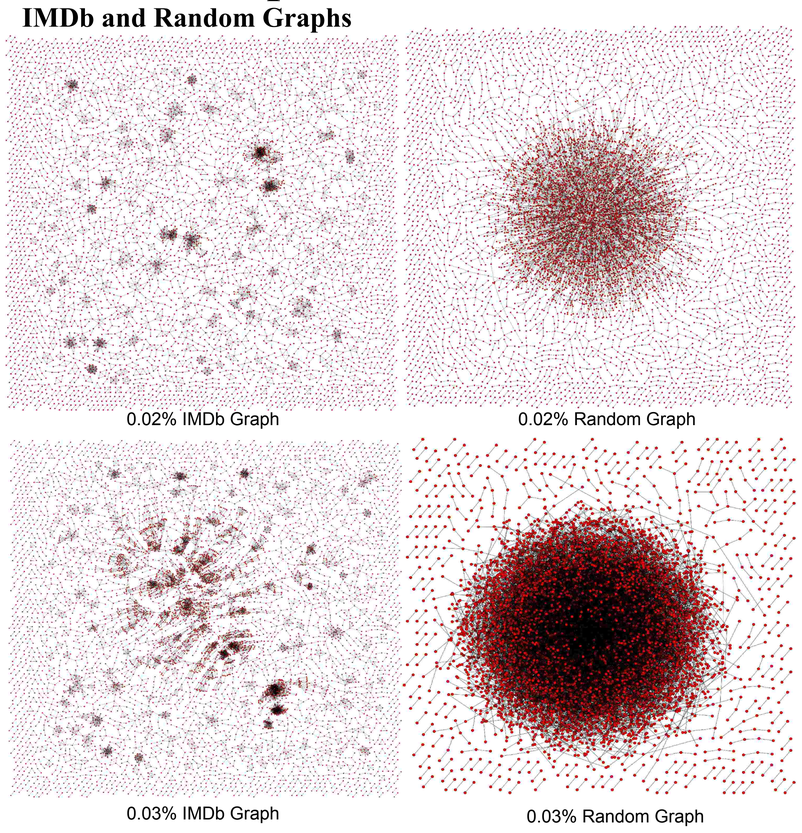
- Sort all the edges according to their distance in non-increasing order.
- Choose top x% of them which may keep the semantic meanings to construct the graphs. Hence, all the graphs are very sparse.
- We selected top 0.003%, 0.02%, 0.08% and 0.5% of all the possible edges to show the results.

Random Graphs

- In the random graph G(n,m) model, a graph is chosen uniformly randomly from the collection of all graphs which have exactly *n* nodes and *m* edges.

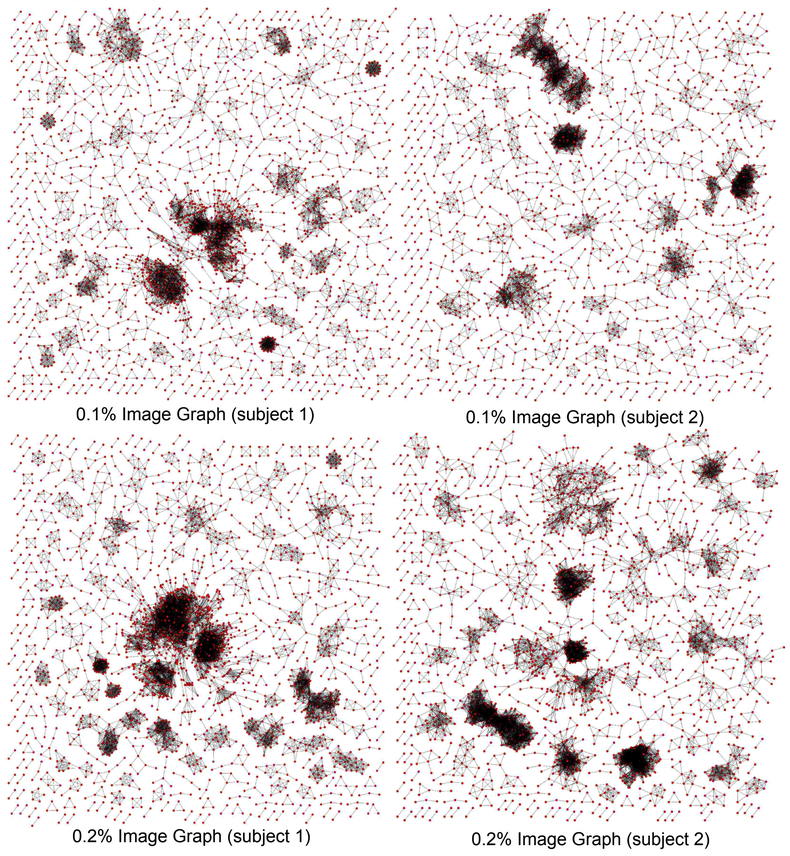
Graphs Visualization

IMDb and Random Graphs



Plots of the graphs above show that even with the same size and same connection densities, the random graphs have different and more simple structures than IMDb graphs.

Image Graphs

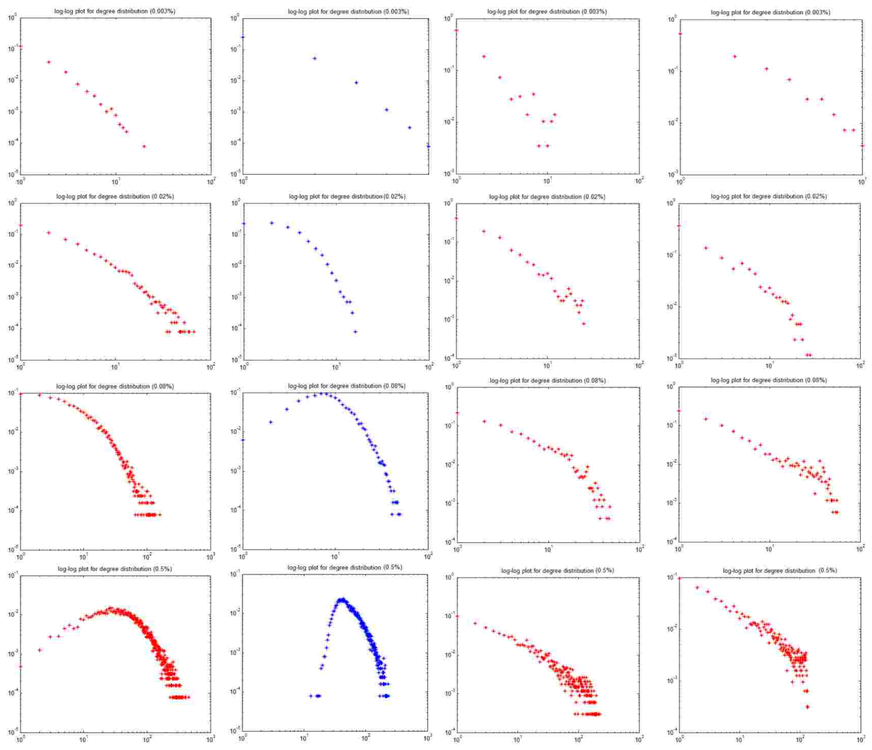


Plots of the graphs above show that the episodic networks, like many other natural networks, have a small-world structure.

Graph-theoretic Analyses of Episodic Graphs

Scale-free properties

- The degree of a vertex in a network is the number of edges incident on or connected to that vertex.
- If there are n nodes in a network and n_k of them have degree k, then the degree distribution p(k) of a network is defined to be $p(k) = n_k/n$.
- A scale-free network is a network whose degree distribution follows a power law.

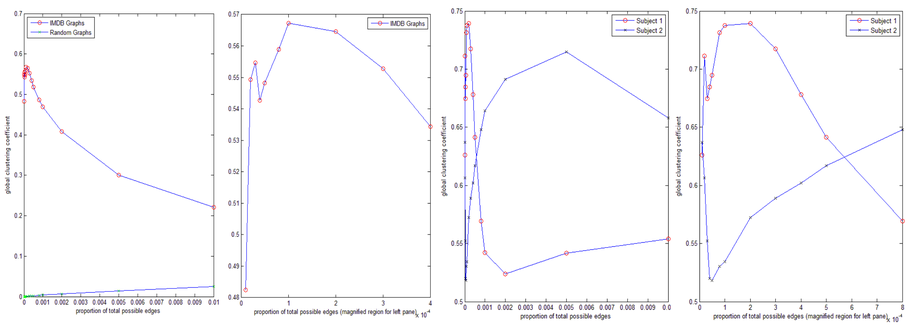


Log-log plots of degree distribution for IMDb graphs and random graphs (four panels on the left side are plots for IMDb graphs, while the panels on the right side are plots for random graphs with the same size and same connection densities.)

Log-log plots of degree distributions for Image graphs (four panels on the left side are plots for subject1's graphs, while the panels on the right side are plots for subject2' s graphs.)

The panels shown from the top to the bottom are log-log plots of the degree distributions with edges proportions from 0.003%, 0.02%, 0.08% and 0.5%.

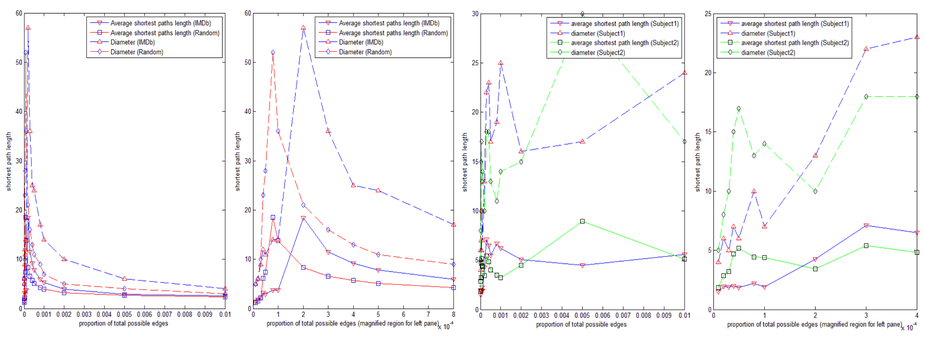
Global clustering coefficient



Global clustering coefficient for IMDb graphs (the right panel is the magnified plot for the coefficient curve of IMDb graphs at the early stage)

Global clustering coefficient for image graphs (the right panel is the magnified plot for the image graphs' clustering coefficient curves at the early stage)

Shortest path lengths



Average shortest path lengths for IMDb graphs(the right panel is the magnified plot for the path length curve of IMDb graphs at the early stage)

Average shortest path lengths for image graphs (the right panel is the magnified plot for the image graphs' path length curves at the early stage)

Conclusion and Discussion

- We show that episodic networks, like many other natural networks, have a small-world structure which is characterized by a high clustering coefficient and a short average shortest path length.
- We don't observe a scale-free property from the log-log plots of degree distributions for these two episodic graphs.