Expansion Properties of Large Social Graphs

Fragkiskos D. Malliaros and Vasileios Megalooikonomou

Department of Computer Engineering and Informatics
University of Patras, Greece

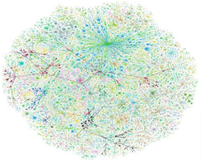
2nd International Workshop on Social Networks and Social Media Mining on the Web (SNSMW)

Hong Kong

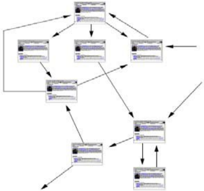
April 22, 2011

Outline

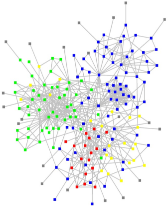
Networks are Everywhere



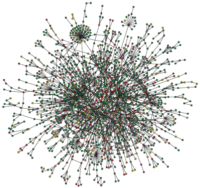
(a) Internet



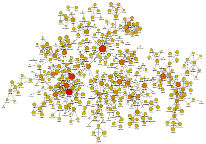
(b) World Wide Web



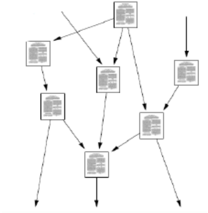
(c) Email network



(d) Protein interactions



(e) Collaboration network



(f) Citation network

Social Media and Networks

- Online social networks and social media
- Easily accessible network data at large scale
- Opportunity to scale up observations
- Large amounts of data raise new questions

Motivating Questions

- What can we say about the structure, the organization and the evolution of networks?
- What tools can we use to study networks?
- **Is there any difference between networks at different scales?**
- How can we utilize the obtained observations in real applications?

Our focus...

- We study the structure of **large scale** social networks
- We explore the **expansion properties** of these networks
- Comparison with known results from previous studies on small graphs

Motivating Questions

- What can we say about the structure, the organization and the evolution of networks?
- What tools can we use to study networks?
- **Is there any difference between networks at different scales?**
- How can we utilize the obtained observations in real applications?

Our focus...

- We study the structure of **large scale** social networks
- We explore the **expansion properties** of these networks
- Comparison with known results from previous studies on small graphs

Motivating Questions

- What can we say about the structure, the organization and the evolution of networks?
- What tools can we use to study networks?
- **Is there any difference between networks at different scales?**
- How can we utilize the obtained observations in real applications?

Our focus...

- We study the structure of **large scale** social networks
- We explore the **expansion properties** of these networks
- Comparison with known results from previous studies on small graphs

Motivating Questions

- What can we say about the structure, the organization and the evolution of networks?
- What tools can we use to study networks?
- **Is there any difference between networks at different scales?**
- How can we utilize the obtained observations in real applications?

Our focus...

- We study the structure of **large scale** social networks
- We explore the **expansion properties** of these networks
- Comparison with known results from previous studies on small graphs

Motivating Questions

- What can we say about the structure, the organization and the evolution of networks?
- What tools can we use to study networks?
- **Is there any difference between networks at different scales?**
- How can we utilize the obtained observations in real applications?

Our focus...

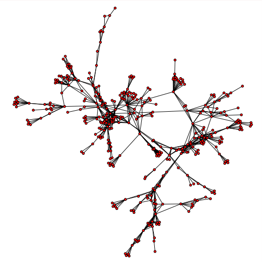
- We study the structure of **large scale** social networks
- We explore the **expansion properties** of these networks
- Comparison with known results from previous studies on small graphs

Study of Structural Properties

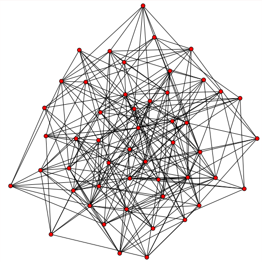
Main Question:

How **large real-world social graphs** look like?

Modular Network

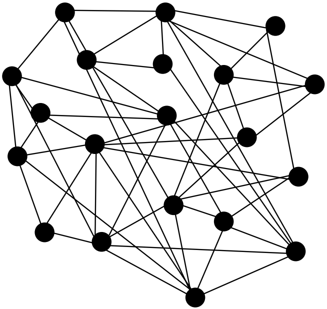


Non Modular Network



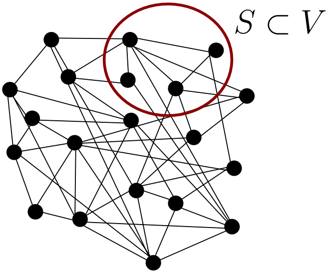
Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**



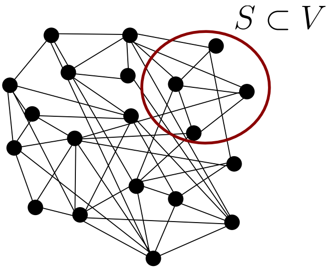
Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**
- Given a graph $G = (V, E)$, the expansion of a subset of nodes S, with $|S| \leq \frac{|V|}{2}$ is $\frac{|N(S)|}{|S|}$



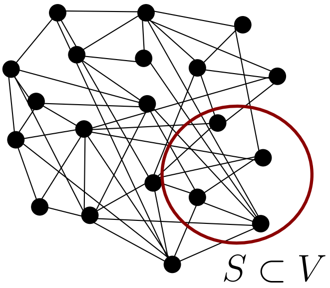
Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**
- Given a graph $G = (V, E)$, the expansion of a subset of nodes S, with $|S| \leq \frac{|V|}{2}$ is $\frac{|N(S)|}{|S|}$



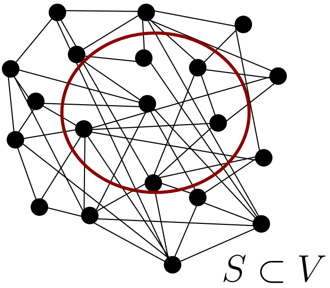
Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**
- Given a graph $G = (V, E)$, the expansion of a subset of nodes S, with $|S| \leq \frac{|V|}{2}$ is $\frac{|N(S)|}{|S|}$



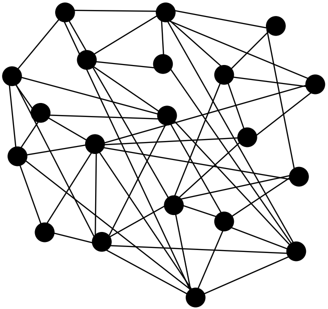
Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**
- Given a graph $G = (V, E)$, the expansion of a subset of nodes S, with $|S| \leq \frac{|V|}{2}$ is $\frac{|N(S)|}{|S|}$



Expansion Properties

- Examine the **expansion properties** or large social graphs
- Graph with **good expansion** properties: simultaneously **sparse** and **highly connected**
- Given a graph $G = (V, E)$, the expansion of a subset of nodes S, with $|S| \leq \frac{|V|}{2}$ is $\frac{|N(S)|}{|S|}$



Expansion factor

$$h(G) = \min_{\left\{S : |S| \leq \frac{|N|}{2}\right\}} \frac{|N(S)|}{|S|}$$

Importance

- They offer crucial insights about the structure of a graph
- They can inform us about the presence or not of edges which can act as bottlenecks inside the network
  - **Modular structure** or lack thereof
- Large expansion factor implies good expansion properties
  - Any subset of nodes will have a relatively large number of edges with one endpoint in this set
  - **Poor modularity**
- Poor expansibility
  - Impossible to satisfy the constraint for a large neighborhood for every subset of nodes
  - The graphs can be easily separated into disconnected subgraphs with the elimination of a small number of edges

Importance

- They offer crucial insights about the structure of a graph
- They can inform us about the presence or not of edges which can act as bottlenecks inside the network
  - **Modular structure** or lack thereof
- Large expansion factor implies good expansion properties
  - Any subset of nodes will have a relatively large number of edges with one endpoint in this set
  - **Poor modularity**
- Poor expansibility
  - Impossible to satisfy the constraint for a large neighborhood for every subset of nodes
  - The graphs can be easily separated into disconnected subgraphs with the elimination of a small number of edges

Importance

- They offer crucial insights about the structure of a graph
- They can inform us about the presence or not of edges which can act as bottlenecks inside the network
  - **Modular structure** or lack thereof
- Large expansion factor implies good expansion properties
  - Any subset of nodes will have a relatively large number of edges with one endpoint in this set
  - **Poor modularity**
- Poor expansibility
  - Impossible to satisfy the constraint for a large neighborhood for every subset of nodes
  - The graphs can be easily separated into disconnected subgraphs with the elimination of a small number of edges

Importance

- They offer crucial insights about the structure of a graph
- They can inform us about the presence or not of edges which can act as bottlenecks inside the network
  - **Modular structure** or lack thereof
- Large expansion factor implies good expansion properties
  - Any subset of nodes will have a relatively large number of edges with one endpoint in this set
  - **Poor modularity**
- Poor expansibility
  - Impossible to satisfy the constraint for a large neighborhood for every subset of nodes
  - The graphs can be easily separated into disconnected subgraphs with the elimination of a small number of edges

Contributions

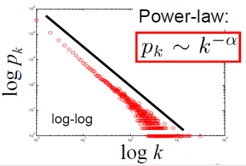
- We measure the expansibility of several **large social graphs**
- What is the expected behavior?
  - They should exhibit **poor** expansion properties
  - They are organized in **communities**
  - **Communities:** groups of nodes with high density of edges within them, and much lower density between different groups
- What really happens?

Related Work (1/2)

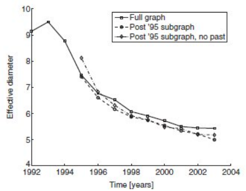
Properties of Networks

- Power-law degree distribution **[Faloutsos et al., 1999]**
- Small diameter **[Albert et al., 1999]**
- Triangle power-law **[Tsourakakis, 2008]**
- Densification power-law **[Leskovec et al., 2005]**
- Shrinking diameter **[Leskovec et al., 2005]**

Degree Distribution



Shrinking diameter



Related Work (2/2)

Expansion Properties

- Estrada studied the expansion properties of several complex networks **[Estrada, 2006]**
- He showed that social networks exhibit poor expansibility
- The work focuses on **small scale networks**
- **Our focus is on large scale social networks**

Measuring Expansion Properties

- Computing the expansion factor
  - Iteration over all possible subsets of nodes with size at most $\frac{|V|}{2}$
  - Computational difficult problem
- Approximation techniques?
- Use the spectrum of the adjacency matrix $\mathbf{A}$
- The expansion factor is closely realated with the **spectral gap** $\Delta\lambda = \lambda_1 - \lambda_2$

Alon - Milman Inequality

$$\frac{\Delta\lambda}{2} \leq h(G) \leq \sqrt{2\lambda_1 \Delta\lambda}$$

Measuring Expansion Properties

- Computing the expansion factor
  - Iteration over all possible subsets of nodes with size at most $\frac{|V|}{2}$
  - Computational difficult problem
- Approximation techniques?
- Use the spectrum of the adjacency matrix $\mathbf{A}$
- The expansion factor is closely realated with the **spectral gap** $\Delta\lambda = \lambda_1 - \lambda_2$

Alon - Milman Inequality

$$\frac{\Delta\lambda}{2} \leq h(G) \leq \sqrt{2\lambda_1 \Delta\lambda}$$

Measuring Expansion Properties

- Computing the expansion factor
  - Iteration over all possible subsets of nodes with size at most $\frac{|V|}{2}$
  - Computational difficult problem
- Approximation techniques?
- Use the spectrum of the adjacency matrix **A**
- The expansion factor is closely realated with the **spectral gap** $\Delta\lambda = \lambda_1 - \lambda_2$

Alon - Milman Inequality

$$\frac{\Delta\lambda}{2} \leq h(G) \leq \sqrt{2\lambda_1 \Delta\lambda}$$

Measuring Expansion Properties

- Computing the expansion factor
  - Iteration over all possible subsets of nodes with size at most $\frac{|V|}{2}$
  - Computational difficult problem
- Approximation techniques?
- Use the spectrum of the adjacency matrix $\mathbf{A}$
- The expansion factor is closely realated with the **spectral gap** $\Delta\lambda = \lambda_1 - \lambda_2$

Alon - Milman Inequality

$$\frac{\Delta\lambda}{2} \leq h(G) \leq \sqrt{2\lambda_1 \Delta\lambda}$$

Measuring Expansion Properties

- Computing the expansion factor
  - Iteration over all possible subsets of nodes with size at most $\frac{|V|}{2}$
  - Computational difficult problem
- Approximation techniques?
- Use the spectrum of the adjacency matrix $\mathbf{A}$
- The expansion factor is closely realated with the **spectral gap** $\Delta\lambda = \lambda_1 - \lambda_2$

Alon - Milman Inequality

$$\frac{\Delta\lambda}{2} \leq h(G) \leq \sqrt{2\lambda_1 \Delta\lambda}$$

First approach

- Large spectral gap implies big expansion factor $\rightarrow$ good expansion properties
- If λ_2 is close enough to λ_1, the spectral gap will be small $\rightarrow$ poor expansibility

Spectral gap based approach

1. Compute the spectral gap
2. If this is large, the graph should have good expansion properties

Problem

How large the spectral gap should be for a graph to have good expansibility?

First approach

- Large spectral gap implies big expansion factor $\rightarrow$ good expansion properties
- If λ_2 is close enough to λ_1, the spectral gap will be small $\rightarrow$ poor expansibility

Spectral gap based approach

1. Compute the spectral gap
2. If this is large, the graph should have good expansion properties

Problem

How large the spectral gap should be for a graph to have good expansibility?

First approach

- Large spectral gap implies big expansion factor $\longrightarrow$ good expansion properties
- If λ_2 is close enough to λ_1, the spectral gap will be small $\longrightarrow$ poor expansibility

Spectral gap based approach

1. Compute the spectral gap
2. If this is large, the graph should have good expansion properties

Problem

How large the spectral gap should be for a graph to have good expansibility?

Our Approach

- Combine the spectral gap with the **subgraph centrality**
- **Subgraph centrality:**
  - It is a centrality measure like degree centrality, betweenness centrality
  - It is based on the number of closed walks starting and ending at node $i \in V$:

$$SC(i) = \sum_{\ell=0}^{\infty} \frac{\mathbf{A}_{ii}^{\ell}}{\ell!}, \ \forall i \in V$$

  - Using techniques from spectral graph theory

$$SC(i) = \sum_{j=1}^{|V|} u_{ij}^2 \sinh(\lambda_j), \ \forall i \in V$$

Characterizing Graphs : the Idea

$$SC(i) = \sum_{j=1}^{|V|} u_{ij}^2 \sinh(\lambda_j), \forall i \in V$$

- Good expansion properties $\rightarrow \lambda_1 \gg \lambda_2$
- In the ***SC*** the first term in the summation will exceed the others
- We can say that

$$SC(i) \approx u_{i1}^2 \sinh(\lambda_1)$$

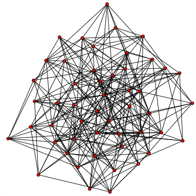
- This implies a **power-law** relationship

$$u_{i1} \propto \sinh^{-1/2}(\lambda_1)\ SC(i)^{1/2}$$

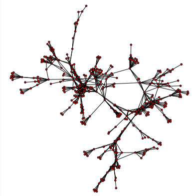
- **Deviations** from this relationship imply absence of good expansibility $\rightarrow \xi(G)$ measure

Example

Random graph

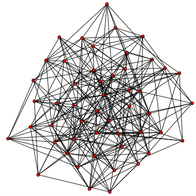


Collaboration graph

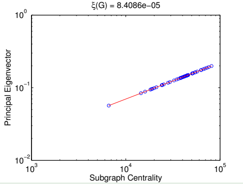


Example

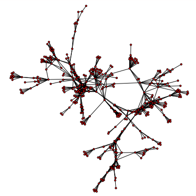
Random graph



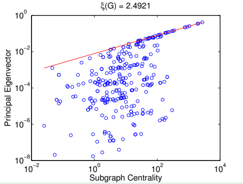
Expansion character



Collaboration graph



Expansion character



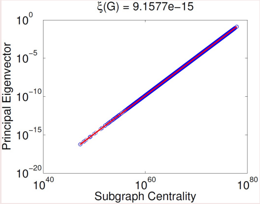
Datasets

Datasets

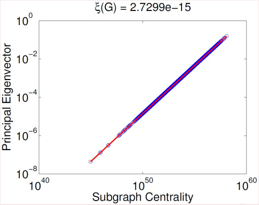
Network	Nodes	Edges
EPINIONS	75,877	405,739
SLASHDOT	77,360	546,487
WIKI-VOTE	7,066	100,736
FACEBOOK	63,392	816,886
YOUTUBE	1,134,890	2,987,624
CA-ASTRO-PH	17,903	197,031
CA-GR-QC	4,158	13,428
CA-HEP-TH	8,638	24,827

Experimental Results (1/2)

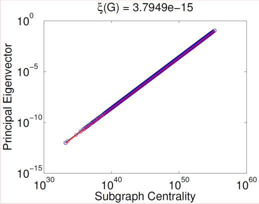
Epinions



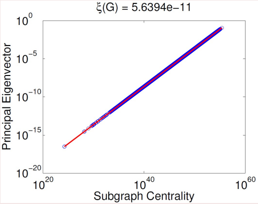
Wiki-Vote



Slashdot

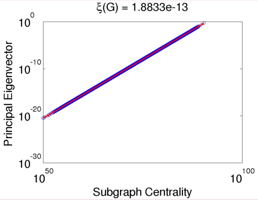


Facebook

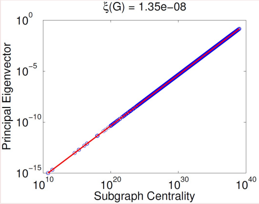


Experimental Results (2/2)

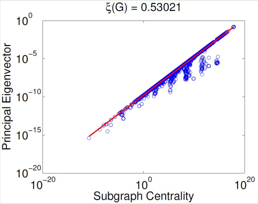
YouTube



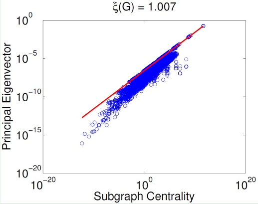
CA-ASTRO-PH



CA-GR-QC



CA-HEP-TH



Interpetation of Observations

- Most of the examined social graphs lack of edges which can act as bottlenecks
- The nodes are not organized based on a clear **modular architecture**
- Absence of well defined clusters which can be easily seperated from the whole graph
- Lack of clusters (communites) with a clear difference between the number of intra-cluster edges and inter-cluster edges
  - Our findings do not imply **absence of communities**

Interpetation of Observations

- Most of the examined social graphs lack of edges which can act as bottlenecks
- The nodes are not organized based on a clear **modular architecture**
- Absence of well defined clusters which can be easily seperated from the whole graph
- Lack of clusters (communites) with a clear difference between the number of intra-cluster edges and inter-cluster edges
  - Our findings do not imply **absence of communities**

Interpetation of Observations

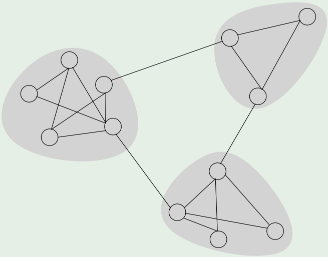
- Most of the examined social graphs lack of edges which can act as bottlenecks
- The nodes are not organized based on a clear **modular architecture**
- Absence of well defined clusters which can be easily seperated from the whole graph
- Lack of clusters (communites) with a clear difference between the number of intra-cluster edges and inter-cluster edges
  - Our findings do not imply **absence of communities**

Interpetation of Observations

- Most of the examined social graphs lack of edges which can act as bottlenecks
- The nodes are not organized based on a clear **modular architecture**
- Absence of well defined clusters which can be easily seperated from the whole graph
- Lack of clusters (communites) with a clear difference between the number of intra-cluster edges and inter-cluster edges
  - Our findings do not imply **absence of communities**

Are the Observations Expected?

Community structure



Social networks

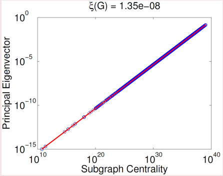
- Social networks $\equiv$ Community structure
- Poor expansion properties
- Experimentaly observed on **small scale social networks**

Possible Explanations

- The expansion properties of **large scale** social graphs are completely different from that of small scale networks
- Mainly due to two reasons:
  1. The **scale** of the network
  2. Social networking and social media applications
     - It is easier for an interaction to be achieved

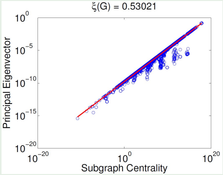
CA-ASTRO-PH

$|V| = 18K, |E| = 200K$



CA-GR-QC

$|V| = 4K, |E| = 13K$



Conclusions

- We measured the expansion properties of several large scale social graphs
- Large scale social graphs, in contrast to small ones, generally exhibit good expansibility
- Structural differences between small and large scale social graphs
- These observations can be possibly utilized in several domains and applications

References I

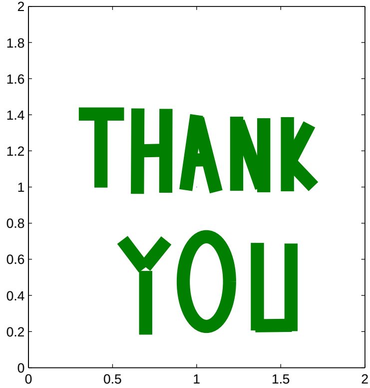
Faloutsos, M., Faloutsos, P., Faloutsos, C.
On power-law relationships of the Internet topology.
In SIGCOMM, pp. 251-262, 1999.

Jure Leskovec, Jon Kleinberg and Christos Faloutsos
Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations
In KDD '05: Proceedings of the 11th International Conference on Knowledge Discovery and Data Mining, 2005.

Albert, R., Jeong, H., Barabasi, A.-L.
Diameter of the world wide web.
Nature, 401:130-131, 1999

Charalampos E. Tsourakakis
Fast Counting of Triangles in Large Real Networks without Counting: Algorithms and Laws.
In ICDM '08: Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 2008.

E. Estrada
Spectral scaling and good expansion properties in complex networks.
Europhys. Lett., 73 (4), pp. 649-655, 2006.


THANK
YOU
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0
0.5
1
1.5
2