ANATOMY OF THE FACEBOOK SOCIAL GRAPH

Source

"*The Anatomy of the Facebook Social Graph*"

By Johan Ugander, Brian Karrer, Lars Backstrom, Cameron Marlow. Arxiv preprint arXiv:1111.4503, 2011

OUTLINE

ANATOMY OF FB SOCIAL GRAPH

1. **Abstract & Introduction**
2. Degree Distribution
3. Path Length
4. Component sizes
5. Clustering coeff. & degeneracy
6. Friends of Friends
7. Exploratory comparison using user traits like age & country.

ABSTRACT: FACEBOOK SOCIAL GRAPH

The paper characterizes the structure of the Facebook social graph using many metrics and tools and discusses various factors:

The number of users and friendships, the degree distribution, path lengths, clustering, and mixing patterns.

DATA

The sample used for this research **contains the entire social network of active members of Facebook in May 2011**, which was 721 million active users.

An **active user** was defined as a user who had logged into Facebook in the last 28 days from the time of measurement in May 2011.
Calculations were performed a Hadoop cluster with 2,250 machines, using the Hadoop/Hive data analysis framework.

INTRODUCTION

- Important since the emergence of SN services, we move from network of documents to a network of people.

- Aims to advance the collective knowledge of social networks & social relationships as embodied in Facebook.

- Active application to the study of everything from bargaining power to psychological health.

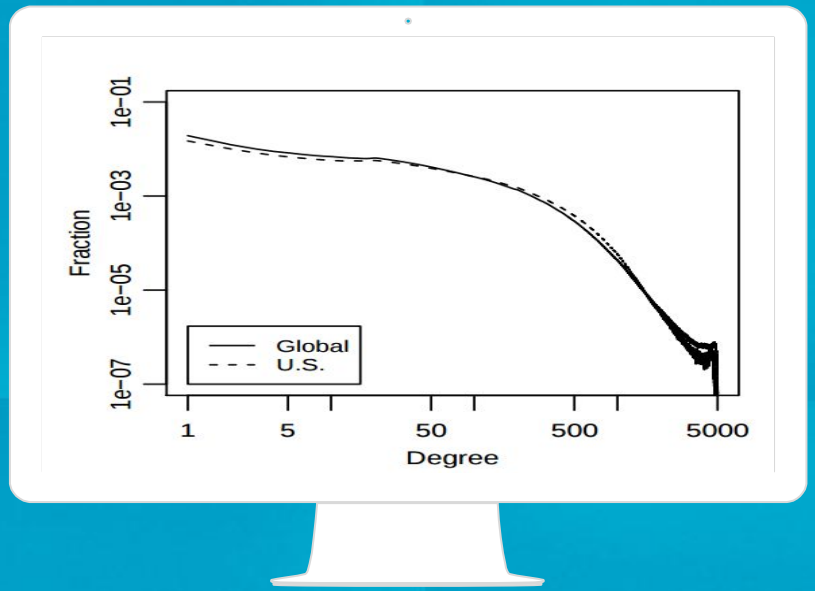
THREE MAIN OBSERVATIONS

1. Presence of a single large connected component
2. Sparse global graph with a dense neighbourhood graph.
3. Phenomenon of “your friends have more friends than you”.

OUTLINE

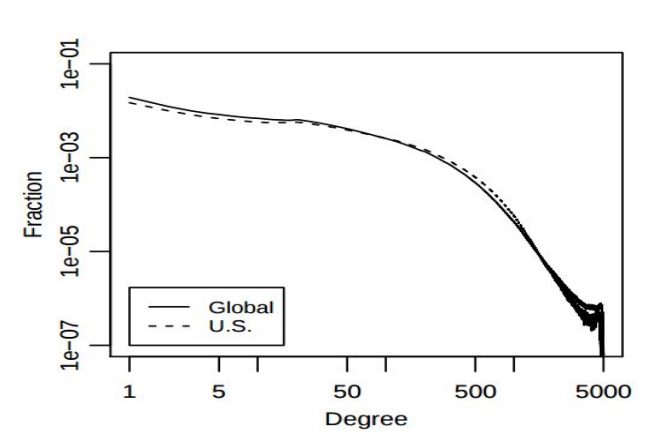
ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. **Degree Distribution**
3. Path Length
4. Component sizes
5. Clustering coeff. & degeneracy
6. Friends of Friends
7. Exploratory comparison using user traits like age & country.



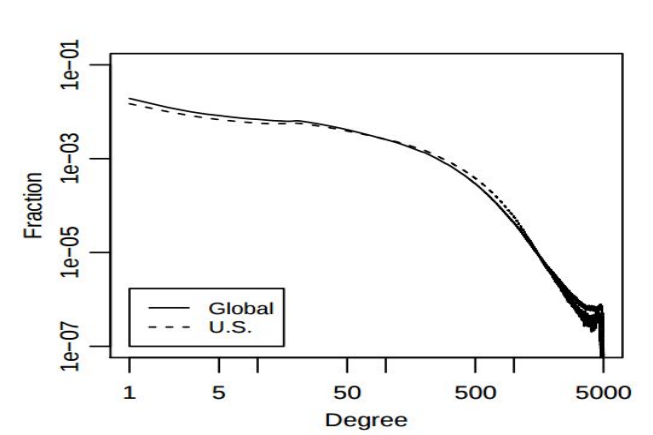
DEGREE DISTRIBUTION

DEGREE DISTRIBUTION



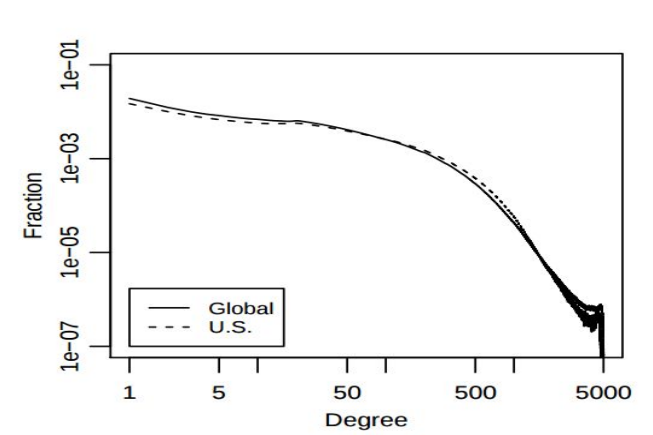
- Does **not** follow the power law.
- Monotonic, except near about 20 friends.
- Curvature of degree distribution.

DEGREE DISTRIBUTION



- Cutoff at 5000 friends(max. friend limit).
- Median: 99

DEGREE DISTRIBUTION

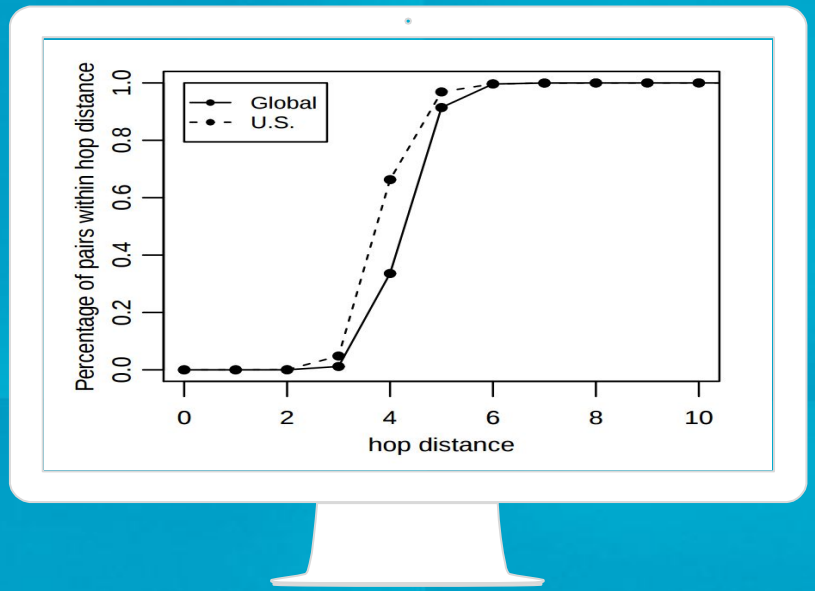


- Most individuals have moderate no. of friends.
- Rightly skewed with a high variance.
- Hubs are also formed but less in no. , “heavy tail”.

OUTLINE

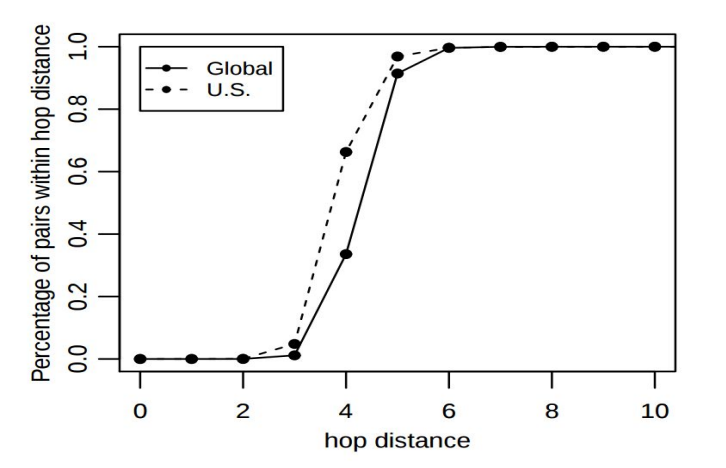
ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. **Path Length**
4. Component sizes
5. Clustering coeff. & degeneracy
6. Friends of Friends
7. Exploratory comparison using user traits like age & country.



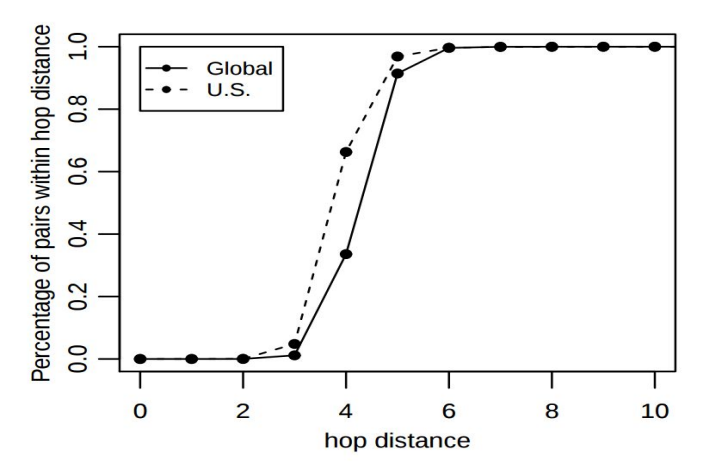
PATH LENGTHS

PATH LENGTHS



- The Facebook graph does not have paths between all pairs of vertices.
- Yet we can plot the neighbourhood function as seen.

PATH LENGTHS

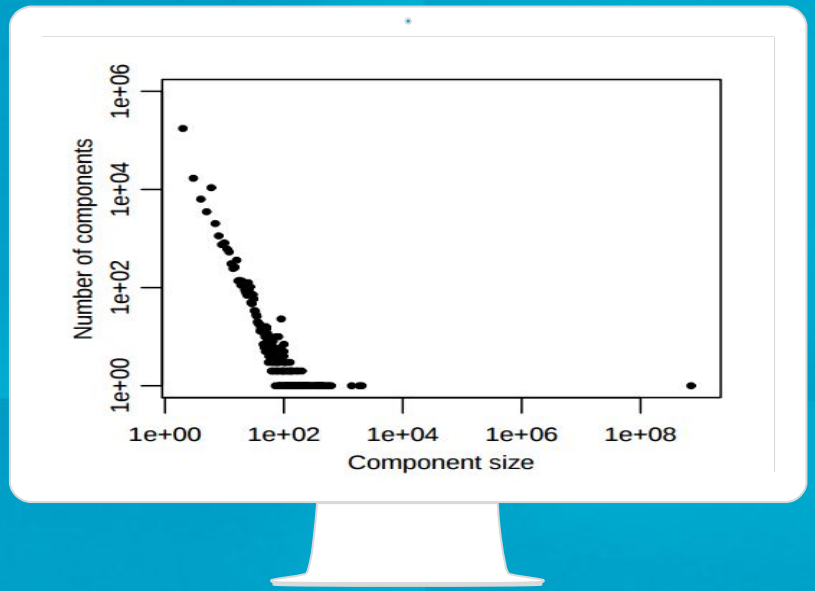


- Average dist. b/w pairs of users: **4.7 for global users. - short path length.**
- Fully 92% of all pairs of Facebook users were within five degrees of separation, and 99.6% were within six degrees.
- Confirms the “six degrees of separation” phenomena.

OUTLINE

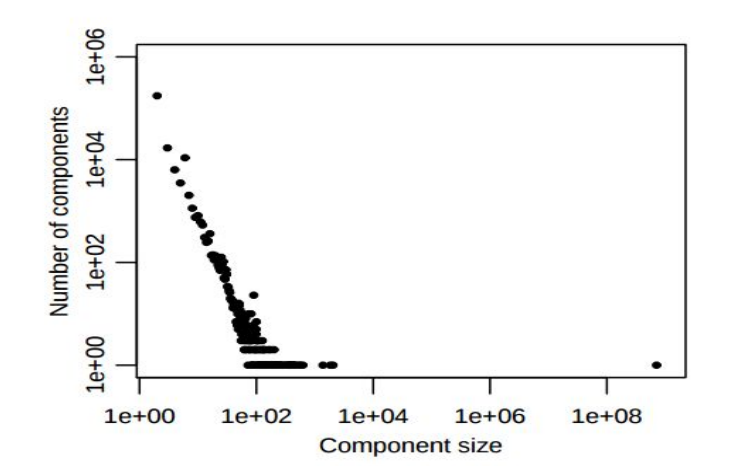
ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. Path Length ✓
4. **Component size**
5. Clustering coeff. & degeneracy
6. Friends of Friends
7. Exploratory comparison using user traits like age & country.



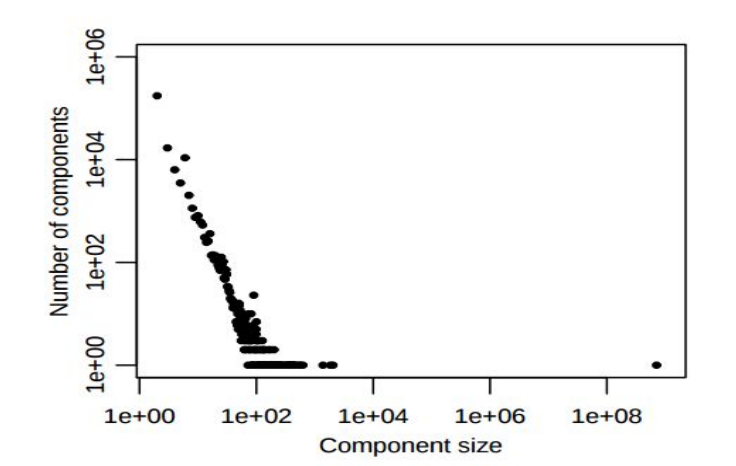
COMPONENT SIZE

COMPONENT SIZE



- Shown is the distribution of component sizes on a log-log scale.
- A **connected component** is a set of individuals for which each pair are connected through a path.
- Most connected components are **extremely small** in size.

COMPONENT SIZE

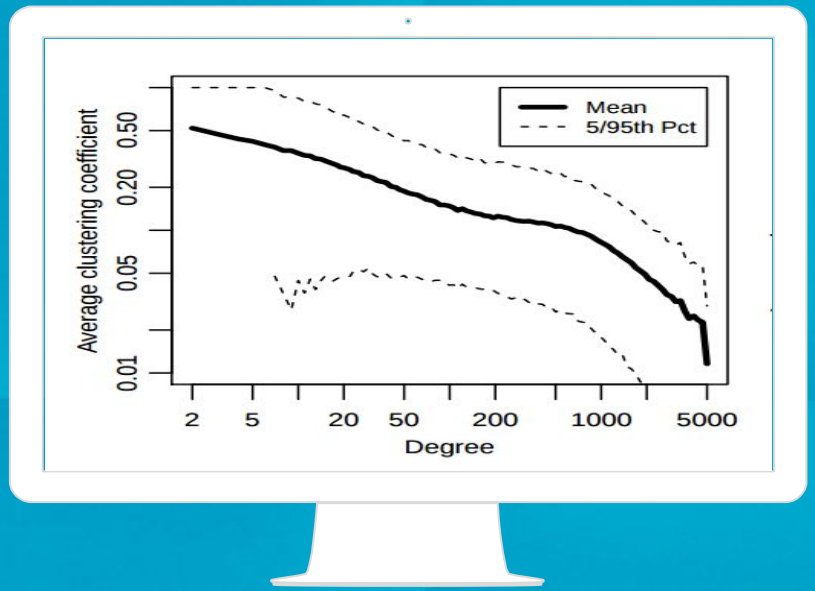


- Largest component has 99.91% of the network.
- Second-largest component has just over 2000 individuals.
- The largest component comprises the vast majority of active Facebook users with at least one friend.

OUTLINE

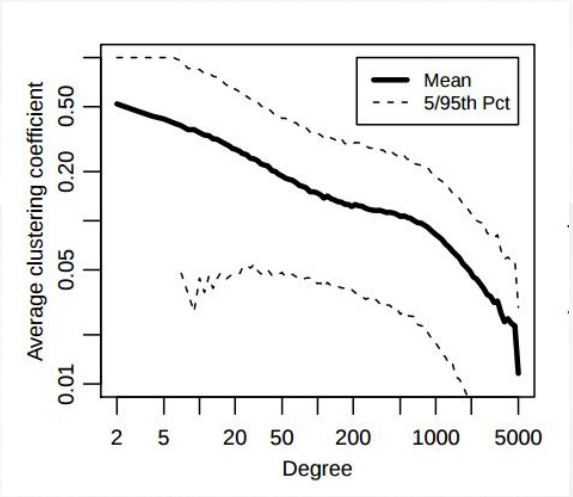
ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. Path Length ✓
4. Component size ✓
5. **Clustering coeff. & degeneracy**
6. Friends of Friends
7. Exploratory comparison using user traits like age & country.



LCC AND DEGENERACY

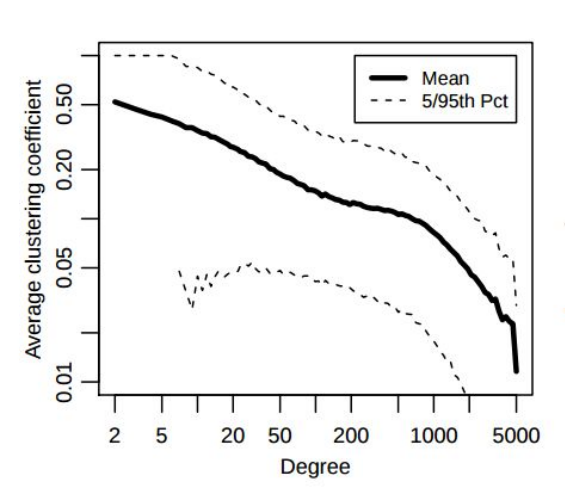
LCC



- **Very large** regardless of the degree, compared to the percentage of possible friendships in the network as a whole.
- Eg. User with 100 friends, lcc = 0.14 ,i.e. , 14% of all its friend pairs are themselves friends.
- 5 times greater than graph of MSN messenger correspondences.

1: Planetary scale views on a large IM network (Leskovec, Horvitz) WWW

LCC

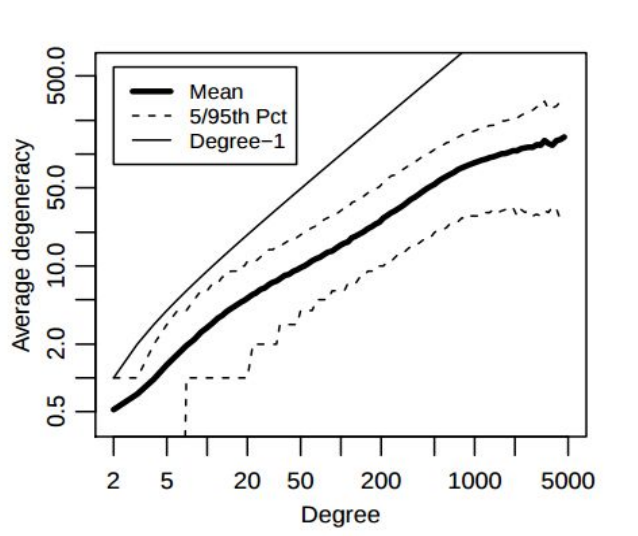


- Analysis also shows that clustering coeff. **decreases** monotonically with the degree, consistent with other studies on MSN messenger.

- LCC **drops rapidly for users with close to 5000 friends**.

- This indicates that these users are more likely using FB for less coherently social purposes, and adding users indiscriminately.

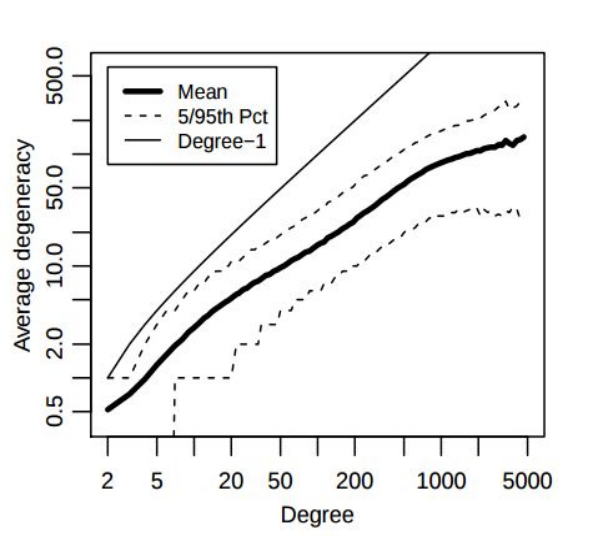
DEGENERACY



- Degeneracy of a graph G is defined as the largest k for which we have a non-empty k-core for G.

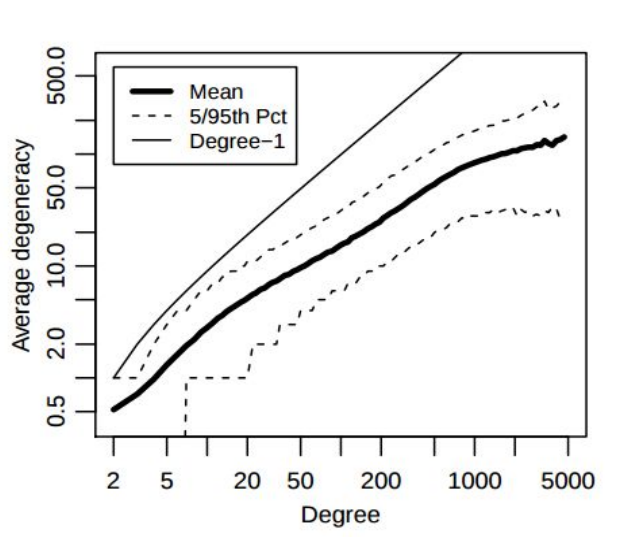
- K-core of a graph G is the subgraph of G which has all nodes whose degree>=k

DEGENERACY



- Provides a robust indication of how tightly-knit a community exists within the graph.
- Avg. degeneracy is an increasing function of user degree.

DEGENERACY

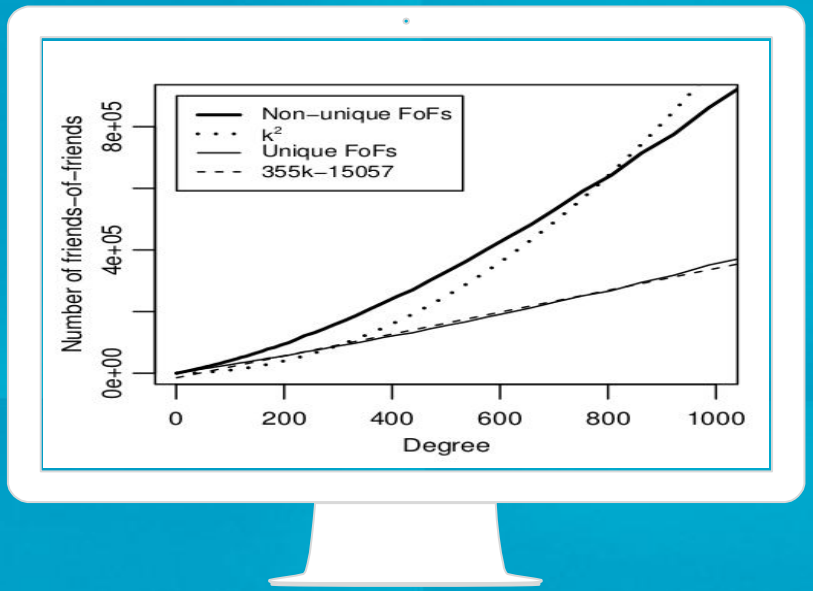


- Quite high avg. degeneracies.
- This is **in contrast** with other research theories. (Eppstein & Stresh 2011)
- Key finding: Even though FB graph is sparse as a whole, **users with sizeable friend counts** do not just have indiscriminate friendships, **rather center around dense network cores**.

OUTLINE

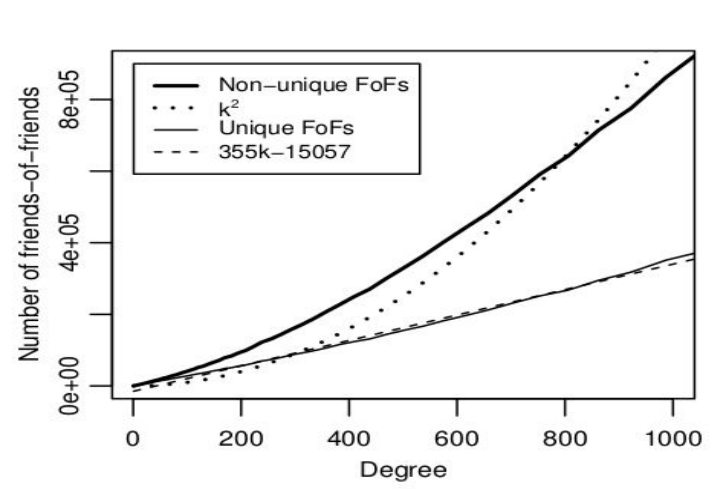
ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. Path Length ✓
4. Component size ✓
5. Clustering coeff. & degeneracy ✓
6. **Friends of Friends**
7. Exploratory comparison using user traits like age & country.



FRIENDS OF FRIENDS

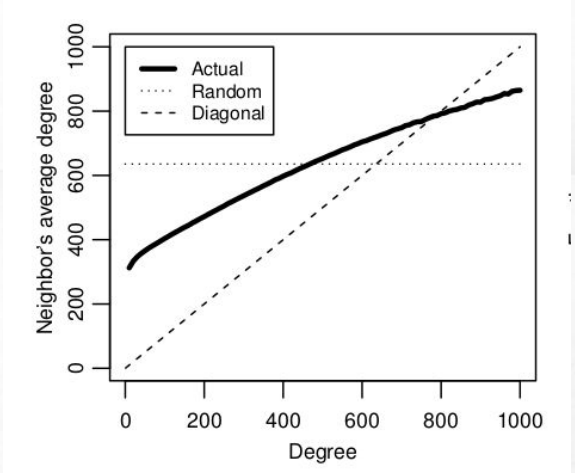
FRIENDS OF FRIENDS



- Computed avg. count of both unique and nonunique FOFs.
- Generally expected: If you have **k** friends, your friends will also have **k** friends, thereby you having at least k^2 non-unique FOFs.
- Turns out, it is **just linear** whereas the non-unique FOFs lies above k^2**.**
- Related to a principle **"your friends have more friends than you."**

FRIENDS OF FRIENDS

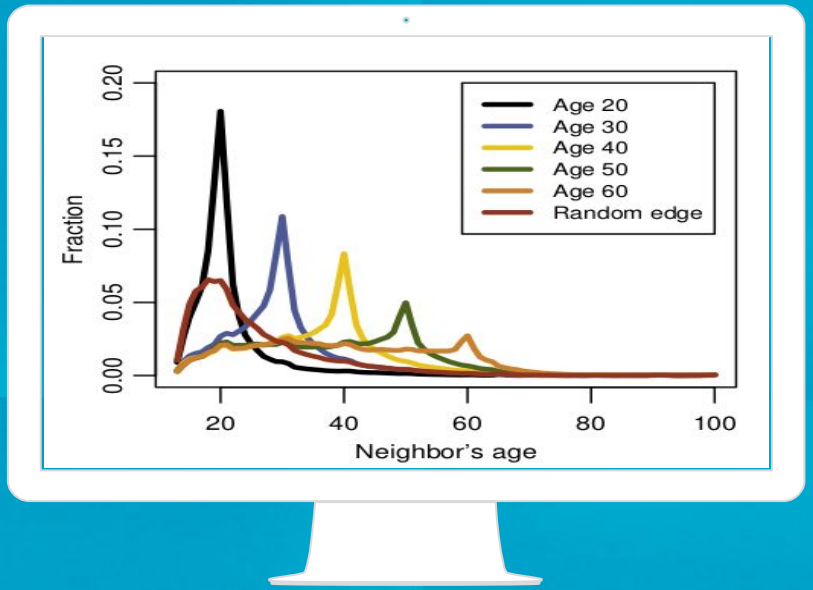
- Avg. no. of friends as shown on the right.
- Until you have 700 friends, your (avg.) neighbor has more friends than you.
- Confirms the phenomena mentioned earlier.



OUTLINE

ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. Path Length ✓
4. Component size ✓
5. Clustering coeff. & degeneracy ✓
6. Friends of Friends ✓
7. **Exploratory comparison using user traits like age & country.**



OTHER METRICS

OTHER METRICS

We discuss a few other metrics and characterise users according to their AGE and COUNTRY OF ORIGIN.

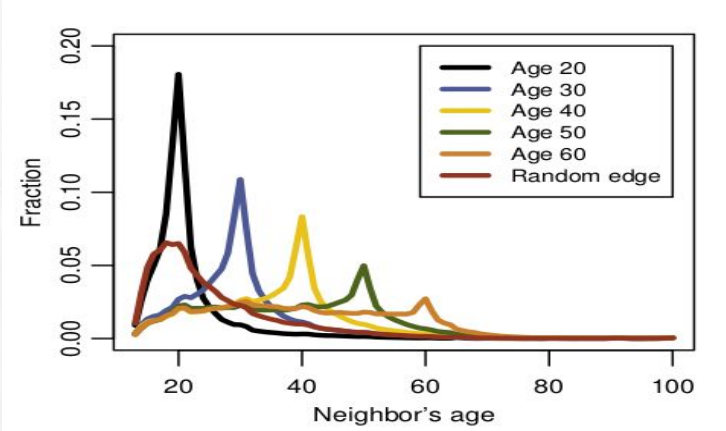
AGE

We start by considering friendship patterns amongst individuals with different ages, and compute the conditional probability p(t ′| t) of selecting a random neighbor of individuals with age t who has age t′ .

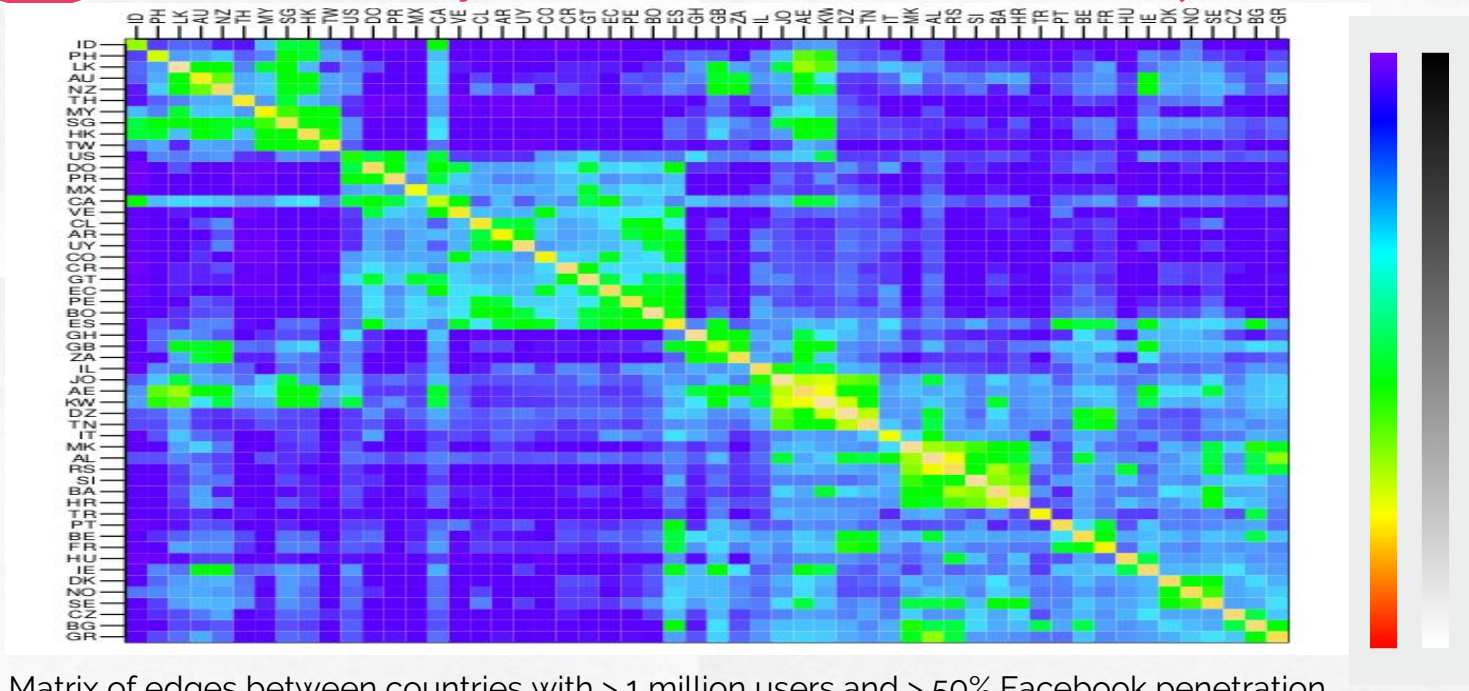
AGE

Unsurprisingly, a random neighbor is most likely to be the same age as you.

Less obviously, the probability of friendship with older individuals falls off rapidly, nearly exponentially, from the mod.

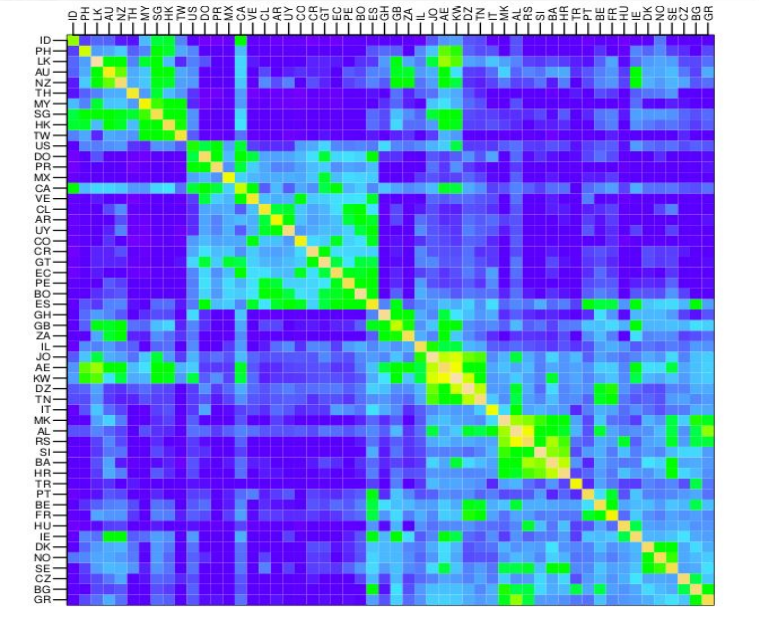


COUNTRY OF ORIGIN



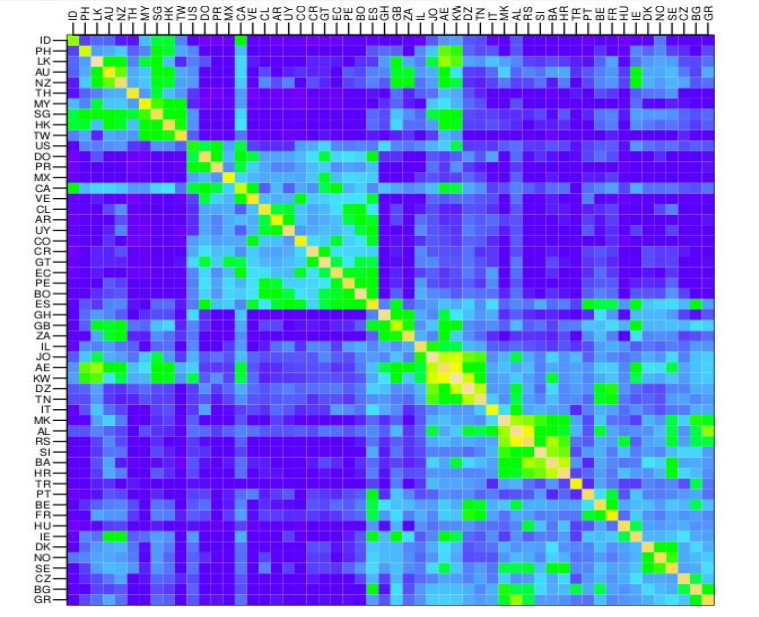
Matrix of edges between countries with > 1 million users and > 50% Facebook penetration.

COUNTRY OF ORIGIN



- The countries fall into groups, the clearly square-like patterns in the matrix, with preferential friendship patterns amongst citizens of different countries.
- Many of the resulting country groupings are intuitive according to geography.
- Eg. It includes the combination of the United Kingdom, Ghana, and South Africa, which may reflect strong historical ties.

COUNTRY OF ORIGIN



- The figure clearly demonstrates that not only are friendships predominantly between users within the same country, but that friendships between countries are also highly modular, and apparently influenced by geography.

OUTLINE

ANATOMY OF FB SOCIAL GRAPH

1. Abstract & Introduction ✓
2. Degree Distribution ✓
3. Path Length ✓
4. Component size ✓
5. Clustering coeff. & degeneracy ✓
6. Friends of Friends ✓
7. Exploratory comparison using user traits like age & country. ✓

RECAP OF THE ANATOMY

Let's review

Power Law

A pure power-law was seen to be inappropriate for the degree distribution of Facebook, although hubs certainly exist.

Friends of friends

Because our friends have more friends than we do, individuals on Facebook have a surprisingly large number of friends-of-friends.

6 Degrees of Separation

The small-world effect and six degrees of separation were then confirmed on a truly global scale.

Clustering

Our friends are highly clustered and our friendships possess dense cores, a phenomena not noticed in smaller social networks. This neighborhood structure has substantial algorithmic implications for graph traversal computations.

Avg. Path Length

It was found to be 4.7, and we interpret this result as indicating that individuals on Facebook have potentially tremendous reach. Shared content only needs to advance a few steps across Facebook's social network to reach a substantial fraction of the world's population.