Real-world Diffusion Phenomena in Complex Networks

Matthieu Latapy, Fabien Tarissan, Daniel Bernardes
LIP6, CNRS and Université Pierre et Marie Curie (UPMC), Paris, France

Information in social networks, viruses in contact networks, files in peer-to-peer networks, etc

Diffusion on a network

A **diffusion trace** is composed of:

1. an underlying graph (the network)
2. chronological data of who transmitted information to whom

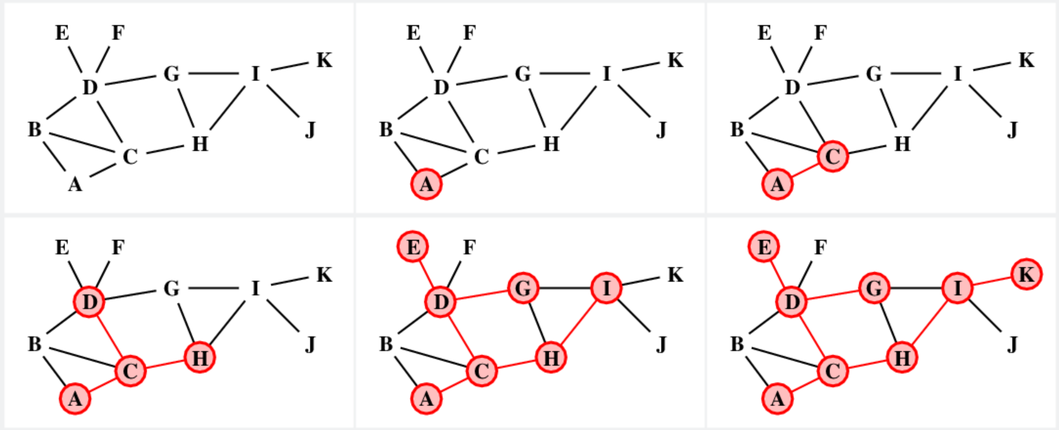


Figure: A diffusion process example.

Modelling

Usual modelling approaches:

- *spreading*: nodes spread the information to a portion of their neighbors
- *adoption*: nodes adopt the information if a portion of their neighbors has it

Validation? Spreading vs adoption? Influence of topology?

In the literature:

empirical data for relatively small-scale diffusion

probabilistic models and numerical simulations

Our goal: study large-scale and real-world diffusion phenomena

Real-world data

In general, the complete diffusion trace is unknown:

Spreading tree
– underlying network?

Nodes reached by the diffusion
– spreading links?

Challenge: obtaining the underlying graph and the spreading events

Our data: trace of file queries to an eDonkey server

file query: { *timestamp, peer id, file id, list of potential providers ids* }

2 days, 5.4 million peers, 2 million files, 212 million queries

Framework

Interest graph: two peers are related if they have a common interest.

Approximation:

two peers are connected if they have requested or provided the same file

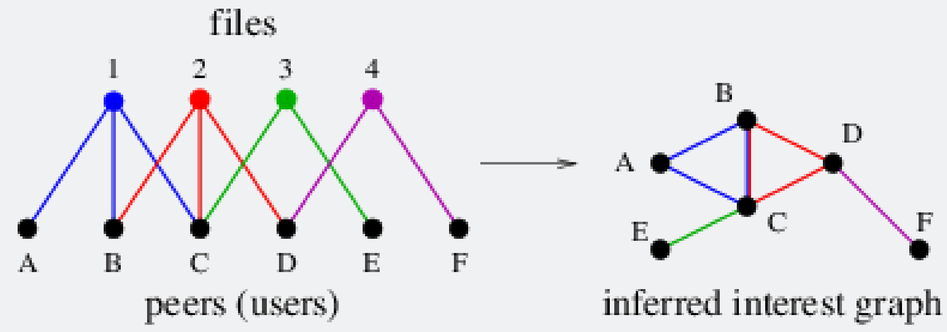


Figure: Interest graph induced by the bipartite graph of file transfers

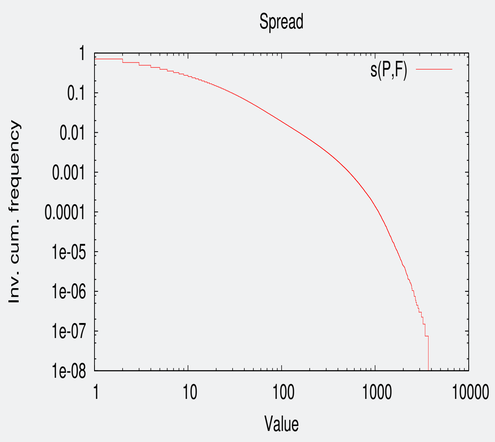
Diffusion of files among peers in the interest graph – key property:

the diffusion takes place in the interest graph

Spreading models

Key parameter: probability to spread a file to a neighbor.

- $\mathbf{s(P,F)}$: *number* of peers to whom peer **P** provides the file **F**.
- $\boldsymbol{\sigma(P,F) = \frac{s(P,F)}{d^o(P)}}$: *fraction* of peers to whom peer **P** provides the file **F**.



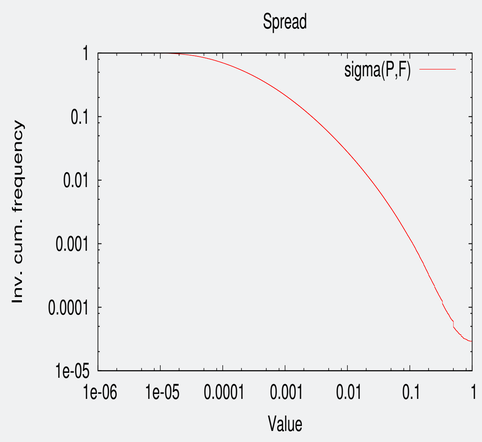
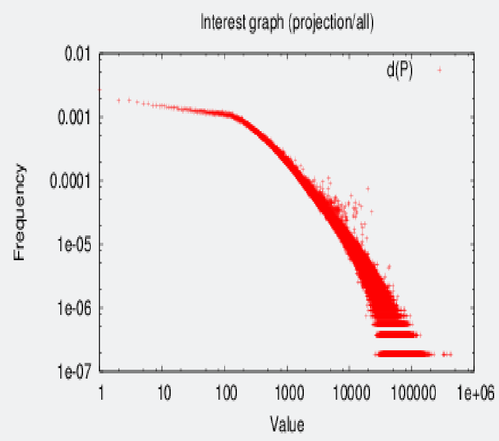


Figure: Inverse cumulative frequency of the spreading parameters $\mathbf{s(P,F)}$ and $\boldsymbol{\sigma(P,F)}$, respectively, for all peers **P** and files **F**.

Result: heterogeneous values (orders of magnitude)

Interest graph properties

Node degrees – median: 827, mean: 3770.36, std. deviation: 8146.85



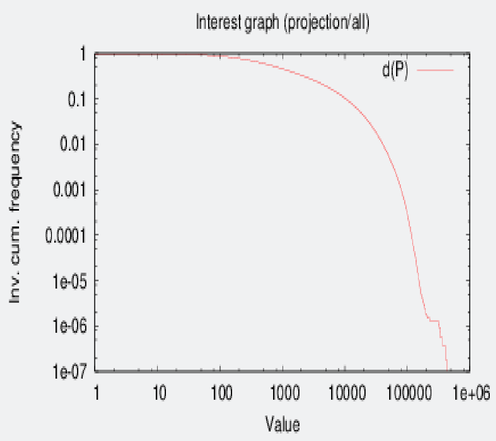


Figure: Frequency and inverse cumulative frequency for the values of node degrees in the interest graph show significantly heterogeneous values (but no power law distribution).

Conclusion

Heterogeneity suggests usual models are inadequate for real-world diffusions