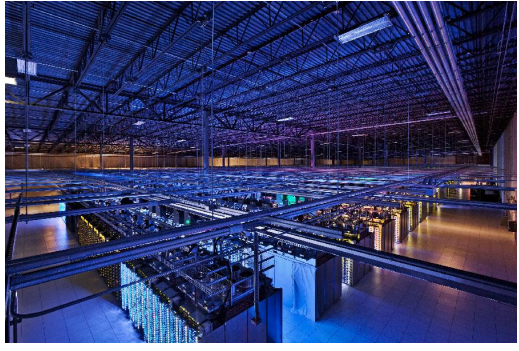
Online social network sites and how to study them

János Török, János Kertész, Yohsuke Murase, Hang-Hyun Jo, Kimmo Kaski

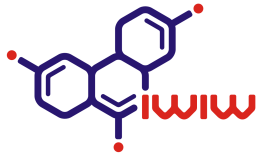
October 22, 2022

Motivation

- Big data: the new buzzword?
- What can we learn from it?
- What are the differences between the real world and data?
- Can we model reality?
- Can we understand people using models motivated by physics?



Data: iWiW



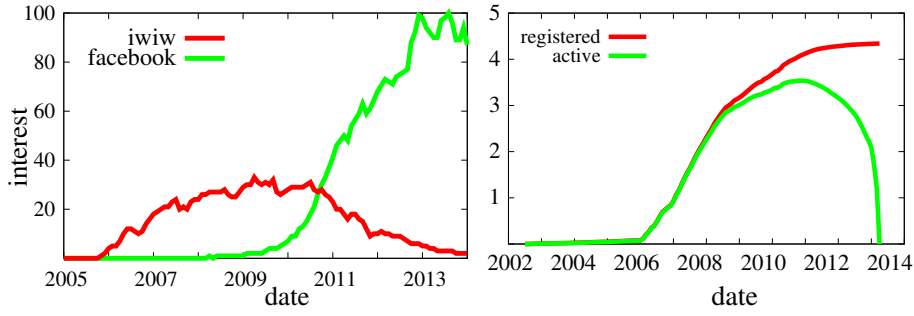
- Established in 2005 (similar to facebook, but more general in the beginning)
- Commercialized in 2006
- Growth by invitation system (linear)
- Top 2-3 site in Hungary for years till 2011
- 35% of the Hungarian population online
- 60% of those with Internet access

Questions

- Why did it fail and how?
- How does ICT data differ from reality?
- What does the data tell us about the society?

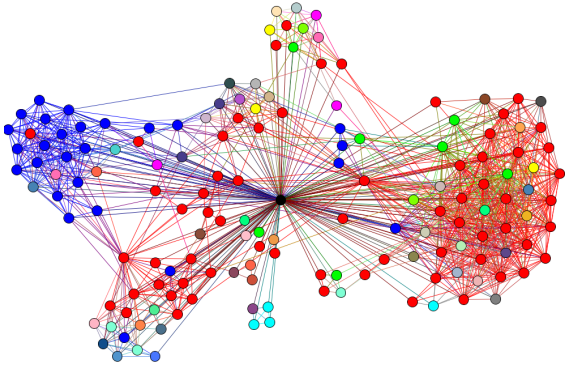
iWiW: Life and death

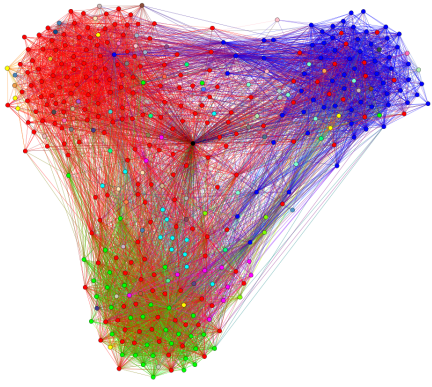
- Linear increase due to limited invitations (one per person in a month or two)
- Growth till end of 2010 (3.5 million active users)
- Stagnation 2011
- Problems 2012 (more than 3 million active users till May 2012)
- Collapse 2013



Egocentric networks

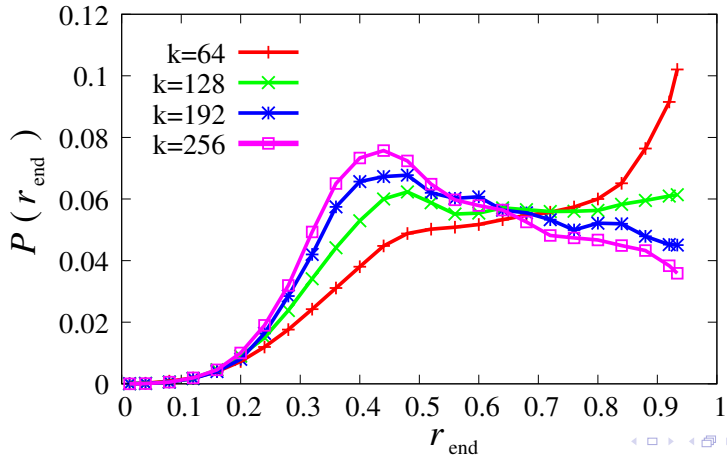
- My egocentric network
- Egocentric network from Becsehely
- Color according to cities





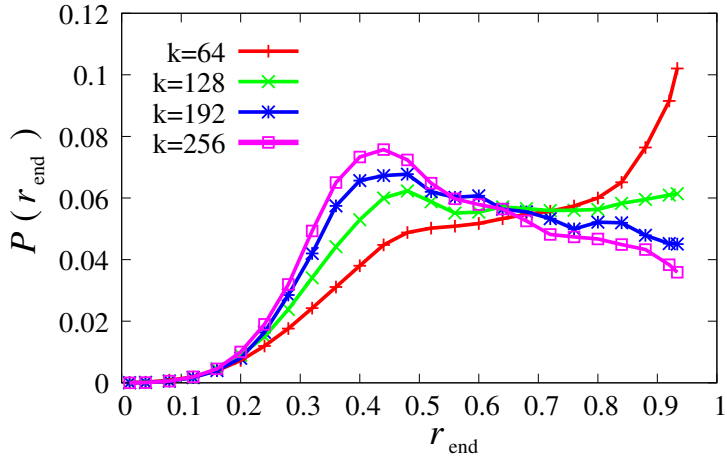
iWiW collapse

- Fraction of active friends at the time of the ego's last login
- k: degree, number of friends on iWiW
- For large k peak at $\sim 0.4-0.45$
- Two week overlap



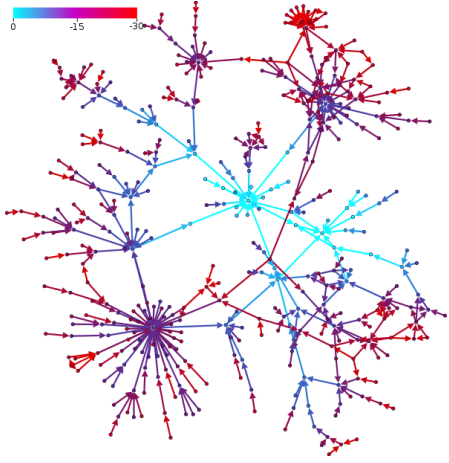
iWiW collapse

- For large k peak at $\sim 0.4-0.45$
- People with limited friends very early
- People with many friends when $\sim 50-60\%$ of their acquaintances left



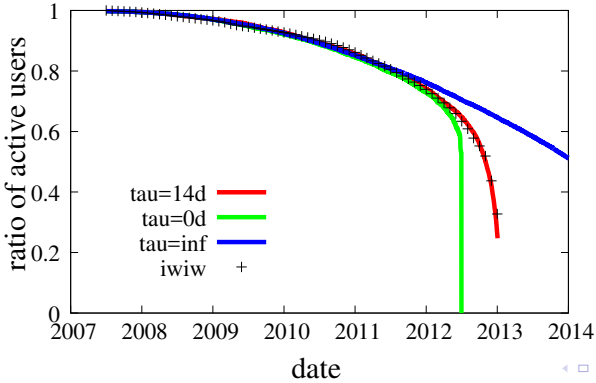
iWiW collapse

- Fraction of active friends at the time of the ego's last login
- k: degree, number of friends on iWiW
- For large k peak at $\sim 0.4-0.45$



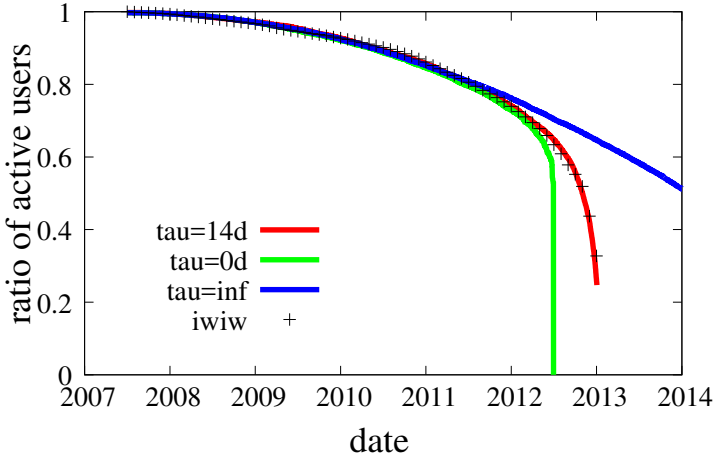
iWiW collapse: cascade model

- Network with average degree $\langle k \rangle$
- There are users that leave with a rate $\gamma = \mu t/\tau$, nodes are chosen with probability proportional to $\sim (k + 25)^{-2}$ (exogenous effects)
- Users for which the ratio of active friends dropped below λ get inactive, but their friends do not realize it immediately
- Users realize the departure of their friend with rate τ



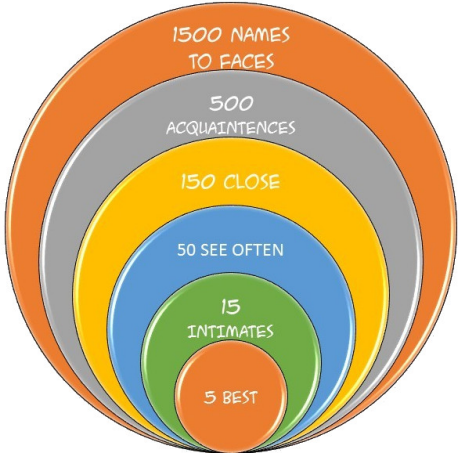
iWiW collapse: model

- Best fit: $\langle k \rangle = 12$, $\tau = 14$ days or $\langle k \rangle = 200$, $\tau = 130$ days
- It takes two weeks to recognize inactive friends



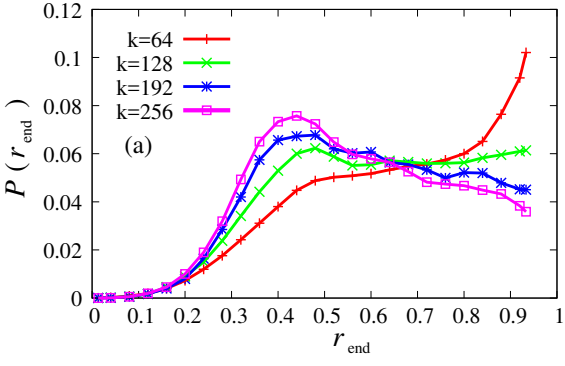
iWiW collapse: model

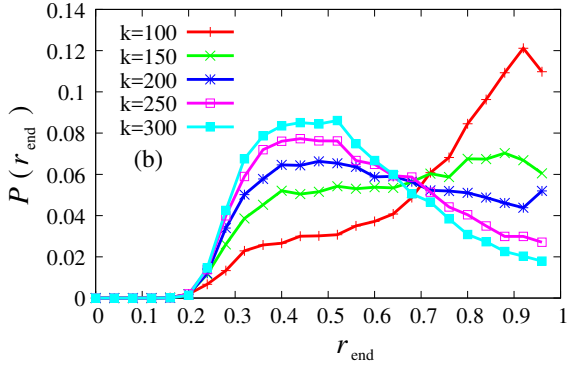
- Best fit: $\langle k \rangle = 12$, $\tau = 14$ days
- It takes two weeks to recognize inactive friends
- It seems, that only the intimate circles matters (Dunbar's circles)



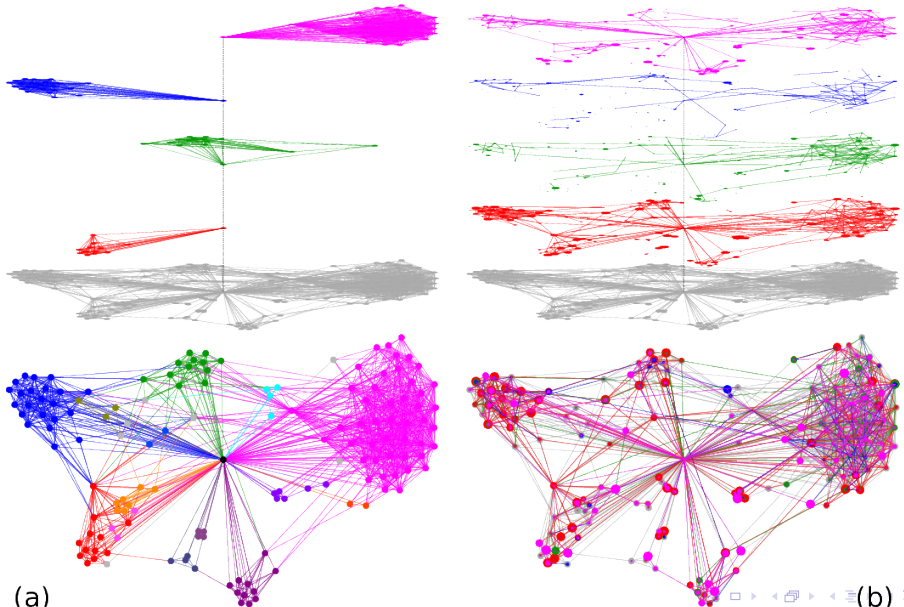
iWiW collapse: model

- Artificial social network
- Cascade model with waiting time
- Degree dependent exogeneous leave

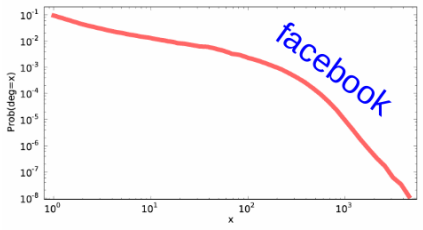


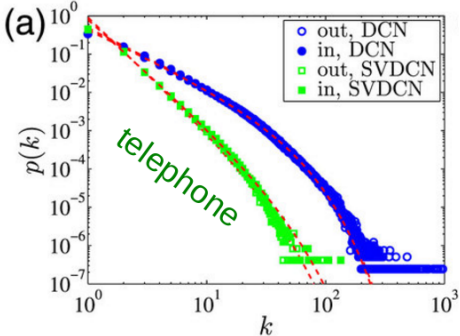


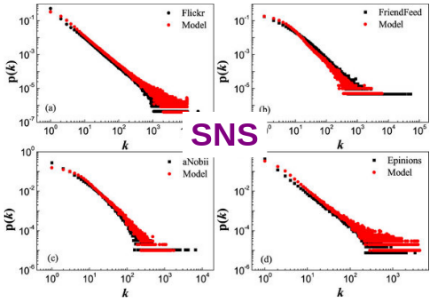
What does ICT data see? Egocentric network:

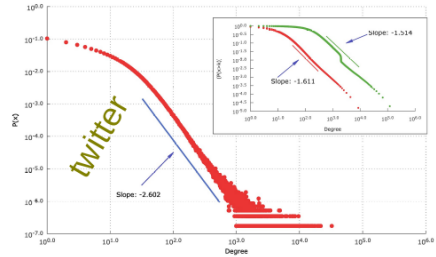


Degree distributions



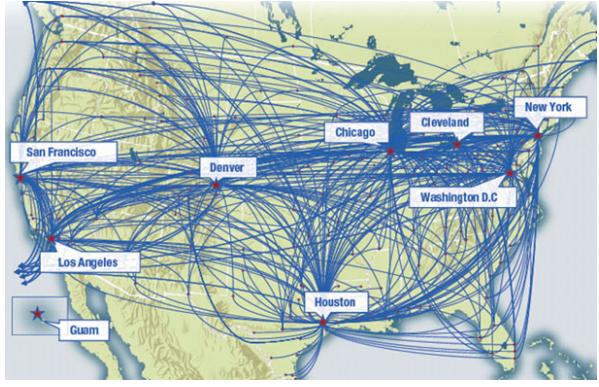




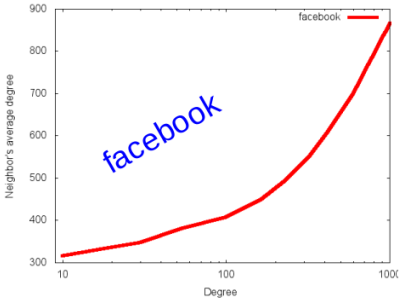


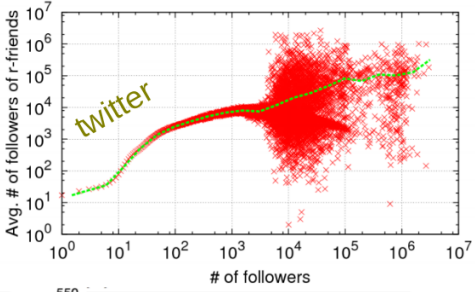
Assortativity

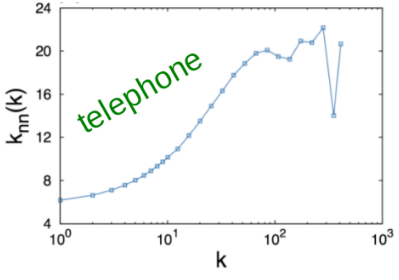
- Average degree of the friends with degree k
- For random networks it is a constant
- For scale free networks it is usually decreasing (disassortative), meaning, that small degree nodes connect to hubs.
- For humans it is believed to be increasing (assortative) extrovert people have extrovert friends.

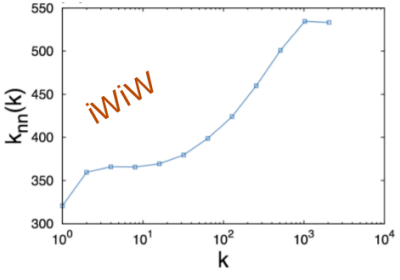


Assortativity



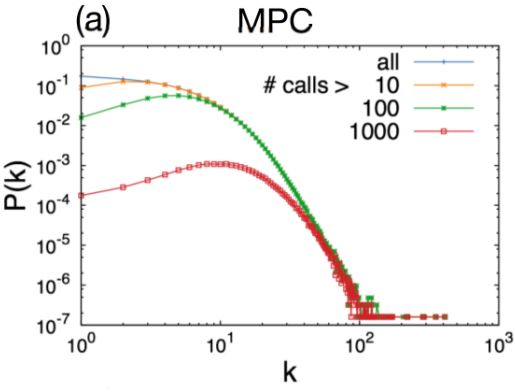


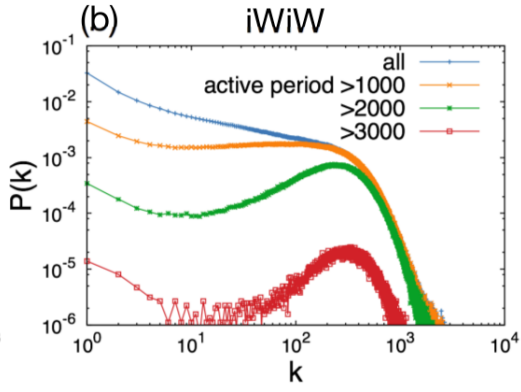




Reality?

- Service usage varies a lot
- Those who use it a lot have more friends on the service
- Maximum at iWiW for experienced users is close to the Dunbar number of acquaintances





Model

- Model: How to choose a communication channel?
  - Which channel do we use to reach a friend?
  - Who's favourite?
  - Least uncomfortable for both!
- A kind of sampling of the underlying social network
- Start from an arbitrary network

Model

- We start from a network (underlying, ground truth)
- Affinity: How much a user likes the channel
- Affinity from exponential distribution

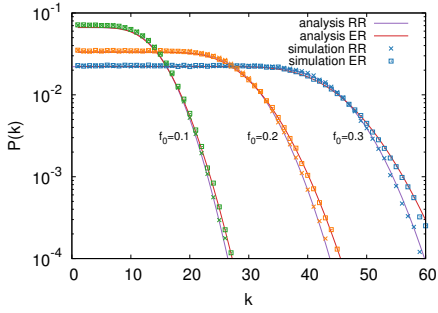
$$P(f) = \frac{1}{f_0} e^{-y/y_0}$$

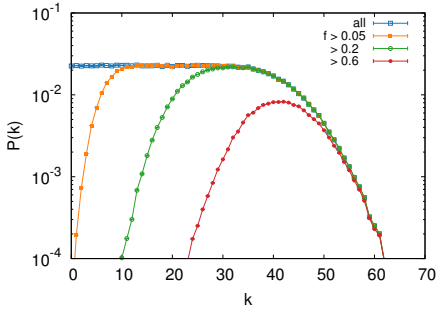
- Assign affinities randomly
- Keep links from the underlying network with probability p_{ij}
- Probability is the minimum of the two

$$p_{ij} = \min(f_i, f_j)$$

Model: Degree distribution

- Model can be solved analytically for Erdős-Rényi and Random Regular network.
- Peak in the degree distribution disappears
- Node with higher affinity have nice peaked degree distribution close to the original





Model: Assortativity