Representation Learning on Networks

Algorithms, Theory, and Applications

Jie Tang
Tsinghua University

Yuxiao Dong
Microsoft Research, Redmond

Tutorial @WWW 2019

A bit about Jie…

- Jie Tang, Professor, Associate Chair of Dept. of Computer Science of Tsinghua University. Interests include social network, data mining, machine learning, knowledge graph.
- I have been visiting scholar at Cornell U. (working with John Hopcroft, Jon Kleinberg), UIUC (working with Jiawei Han), CUHK (with Jeffrey Yu), and HKUST (with Qiong Luo).
- I was awarded with the NSFC for Distinguished Young Scholars, NSFC for Excellent Young Scholars, CCF Young Scientist Award, Newton Advanced Fellowships Award, IBM Innovation Faculty Award, and KDD Service Award.
- Have published more than 200 paper on international conf/journals, including KDD (24), IJCAI/AAAI (24), WWW (8), NIPS/ICML, ACM/IEEE Trans. (27)
- #Citation: 12,466 and *h*-index: 58
- Have a notable system, AMiner.org for academic researcher network analysis. The system has attracted 10 million users from 220 countries/regions.
- **HP:** http://keg.cs.tsinghua.edu.cn/jietang/

A bit about Yuxiao…

- Yuxiao Dong, Senior Applied Scientist at Microsoft Research, Redmond. I received his Ph.D. from University of Notre Dame and has been a visiting scholar at Tsinghua University, U.S. Army Research Lab, and AMiner.org.
- My research focuses on ***data mining, network science, and computational social science,*** with an emphasis on applying computational models to addressing problems in large-scale networked systems, such as the Microsoft Academic Graph (MAG), mobile communication, and online social media.
- My work has been mainly published in KDD and interdisciplinary journals, winning four best paper awards/nominations as well as the ***2017 ACM SIGKDD Doctoral Dissertation Award Honorable Mention.***
- Homepage: https://ericdongyx.github.io/

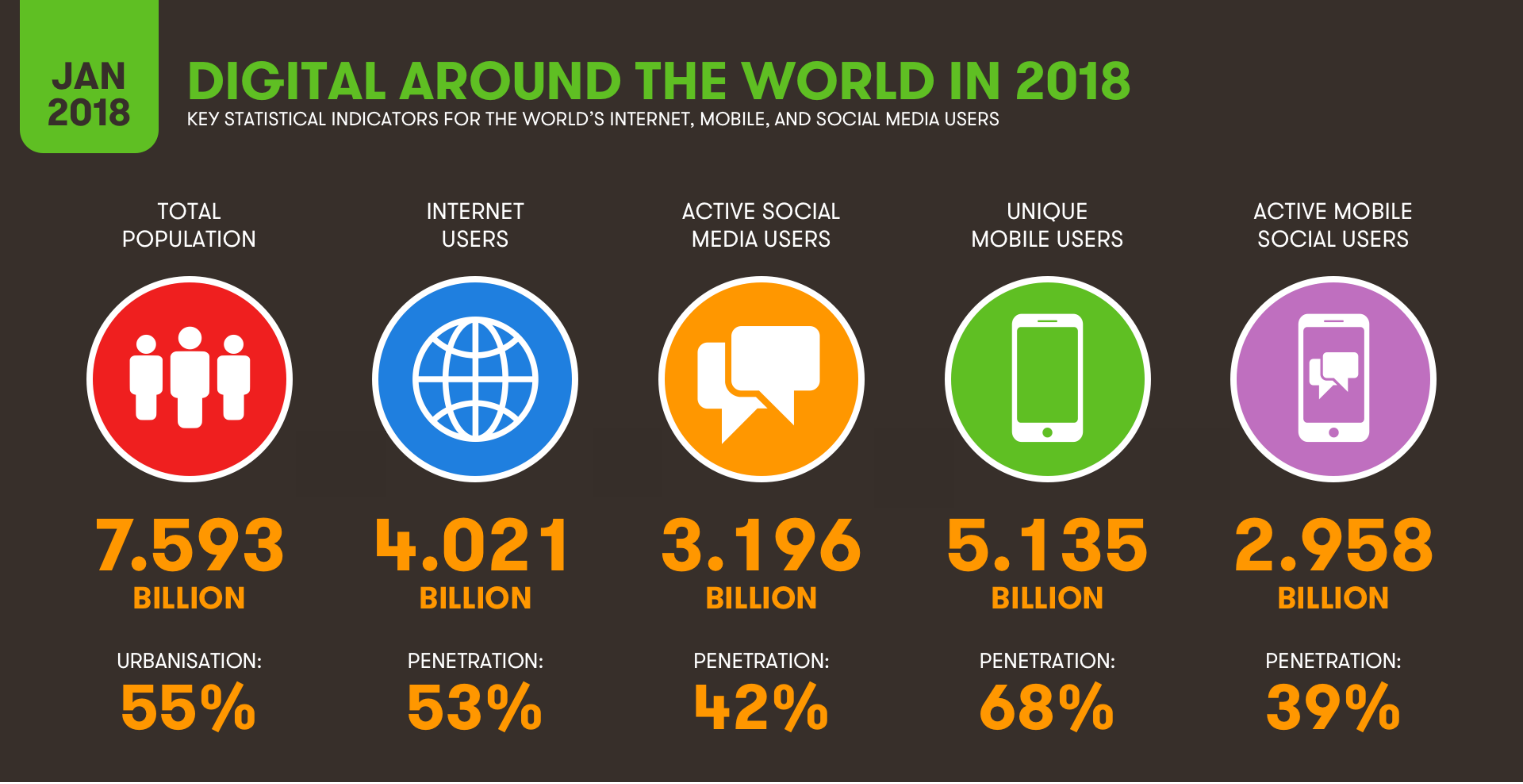


Image Credit: https://wearesocial.com/blog/2018/01/global-digital-report-2018

Physical life coupled with digital world

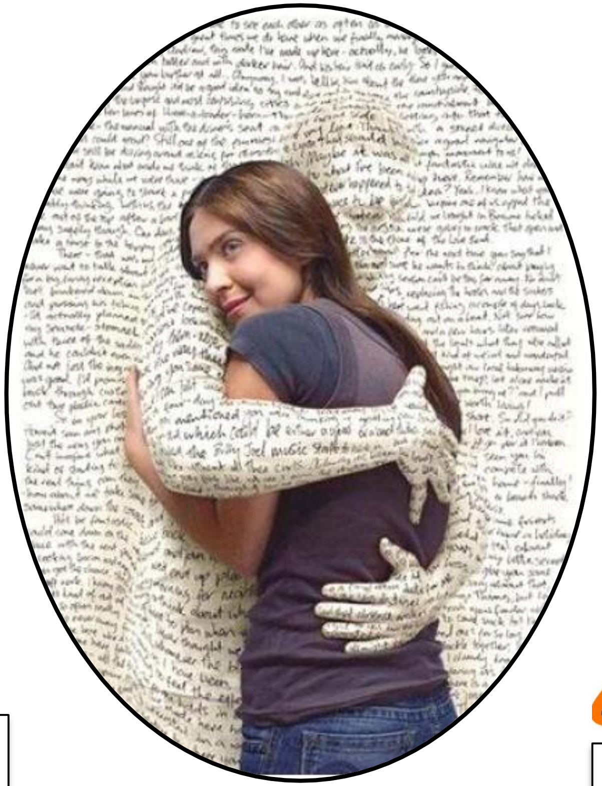
facebook

- FB: **1.47B DAU** & 2.23B MAU
- Instagram: 1B MAU
- 2.5 trillion minutes/month

- 330 million MAU
- Peak: 143K tweets/s

amazon

- 100M Prime users
- $2B on Prime Day



WeChat

- WeChat: **1.04B MAU**
- QQ: 865 million MAU

新浪微博 weibo.com

- 350M MAU
- influencing our daily life

Alibaba Group 阿里巴巴集团

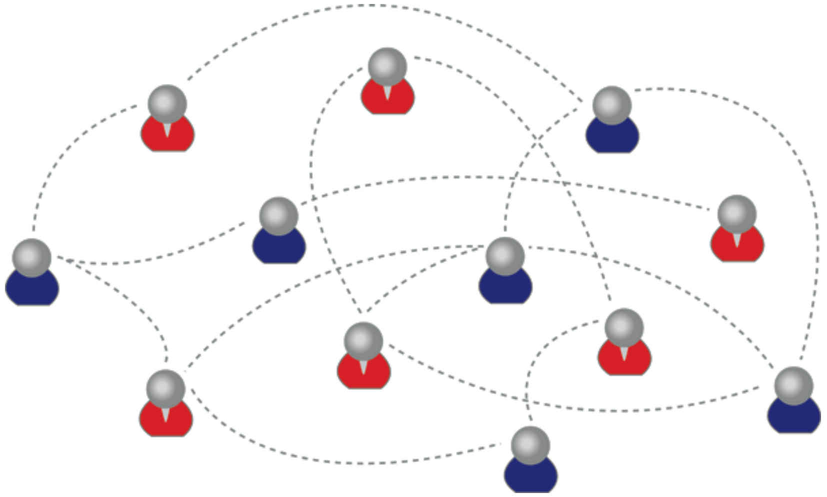
- Alipay: 300M DAU & 777M Trans
- $200B on 11/11/2018 Single’s Day

The era of (digitally) **connected** world

— the world is more **closely** connected than you might think.

A social network is a graph made up of :

– a set of **individuals**, called “nodes”, and
– tied by one or more **interdependency**, such as friendship, called “edges”.

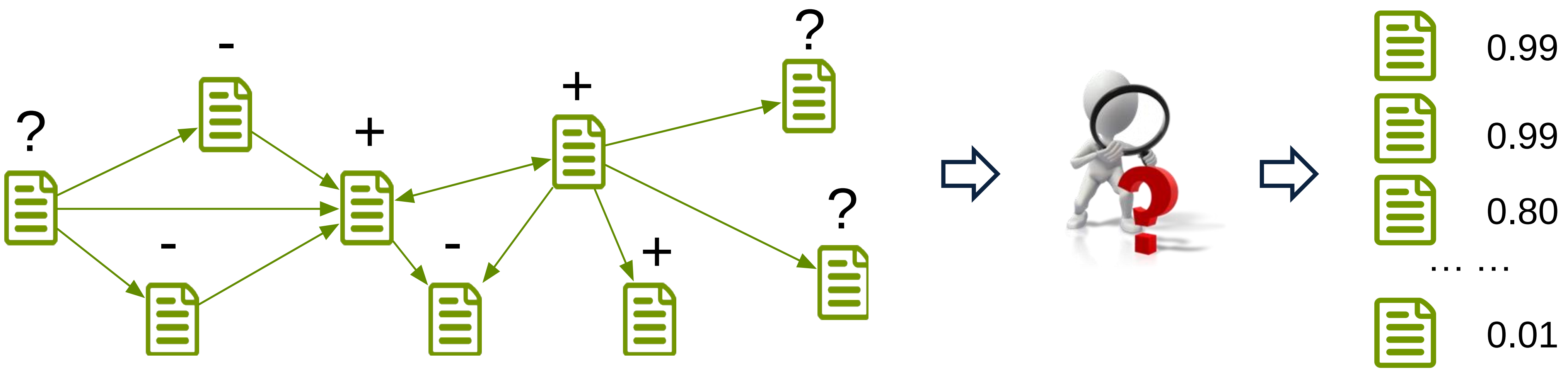


- So, that is ***Social Network***?
- To understand it, we need to trace back…

30 years before…

Web 1.0 = Information Space

- Google's PageRank
- Kleinberg's HITS

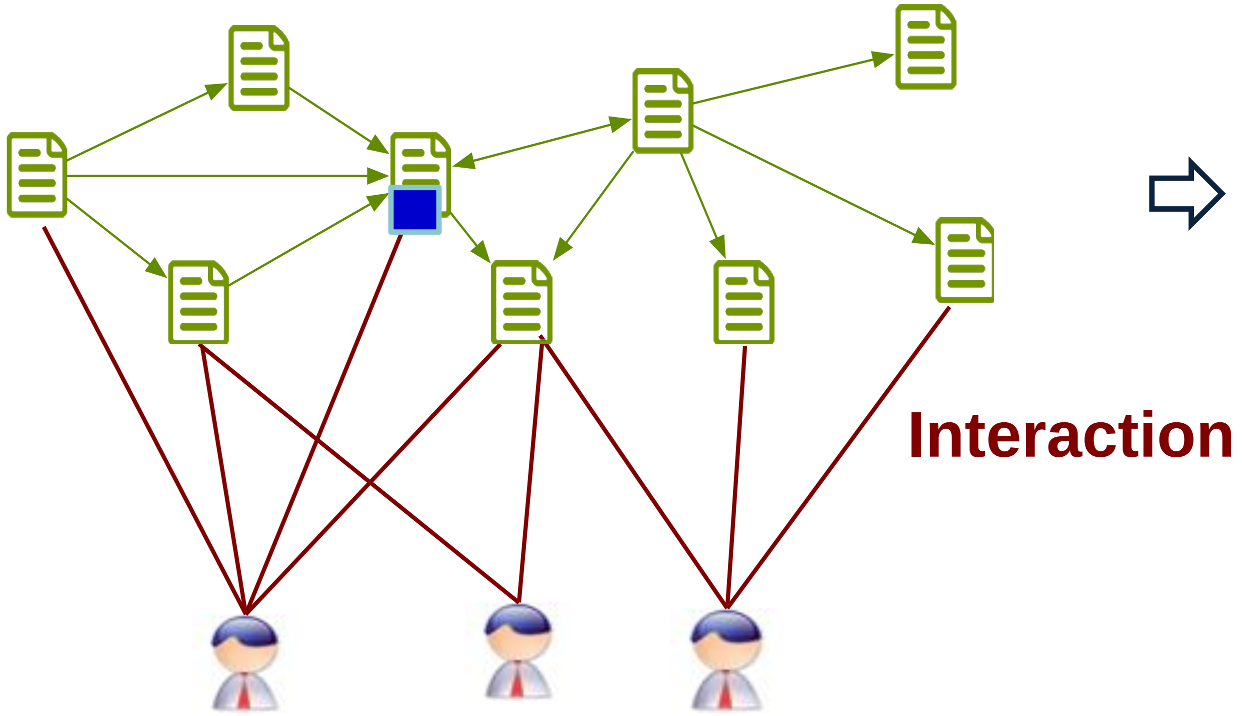


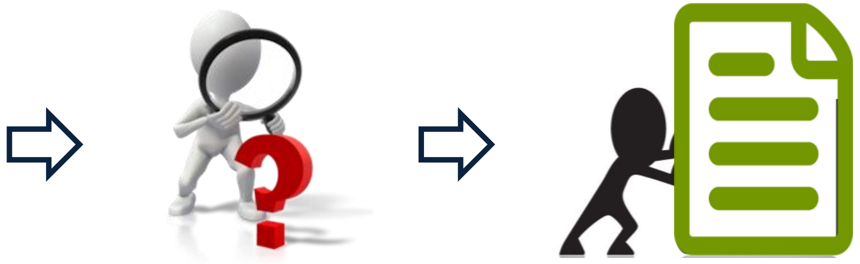
1. L. Page, S. Brin, R. Motwani, & T. Winograd. (1999). The pagerank citation ranking: bringing order to the web. Stanford University.
2. J. M. Kleinberg. Authoritative sources in a hyperlinked environment. Journal of the ACM (**JACM**) 46.5 (1999): 604-632.

20 years before…

Web 2.0 = Info Space + Users

- Personalized recommendation
- Collaborative Filtering

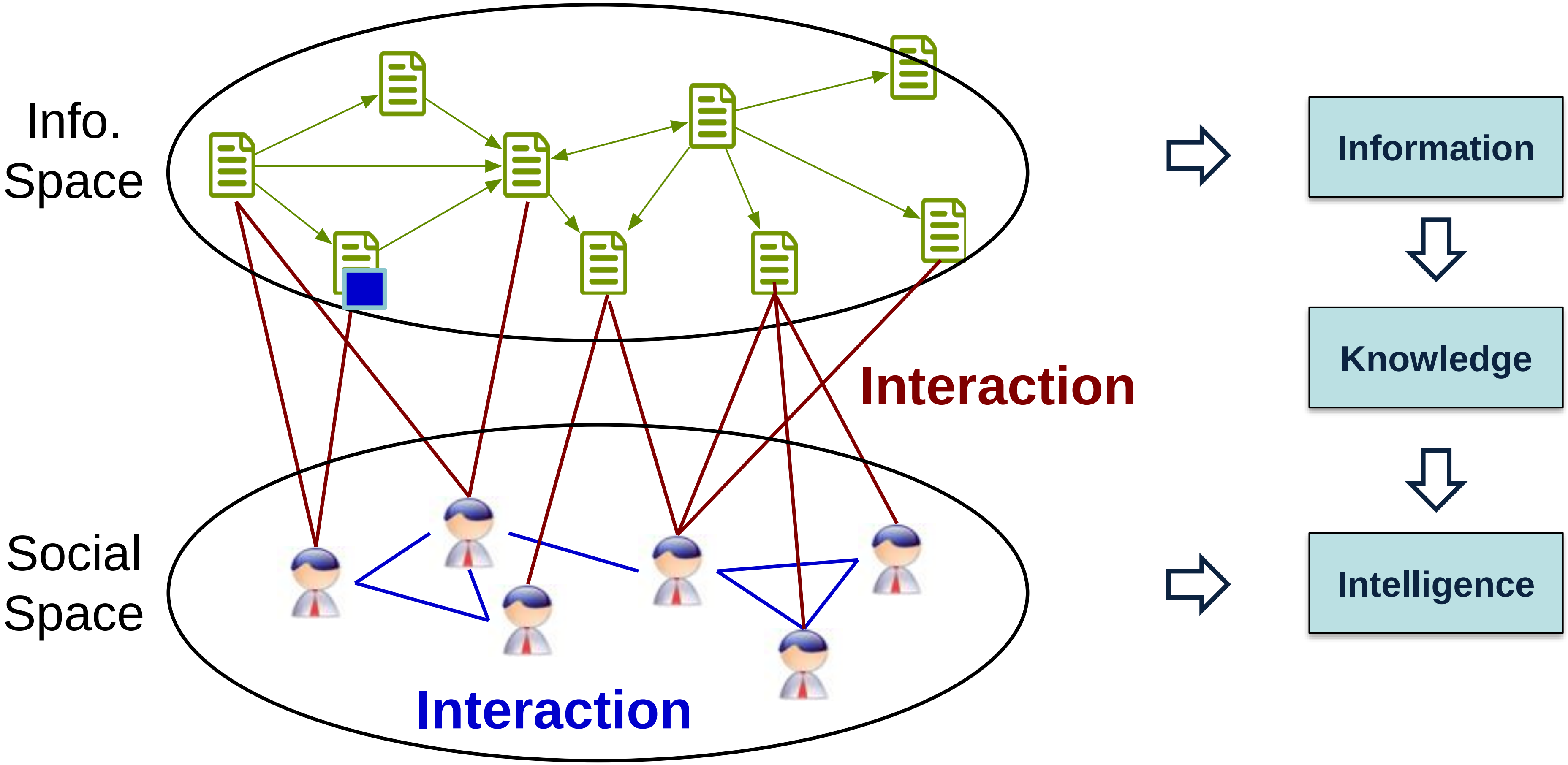




1. P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, & J. Riedl. (1994). GroupLens: an open architecture for collaborative filtering of netnews. ACM **CSCW 1994.**

10 years before…

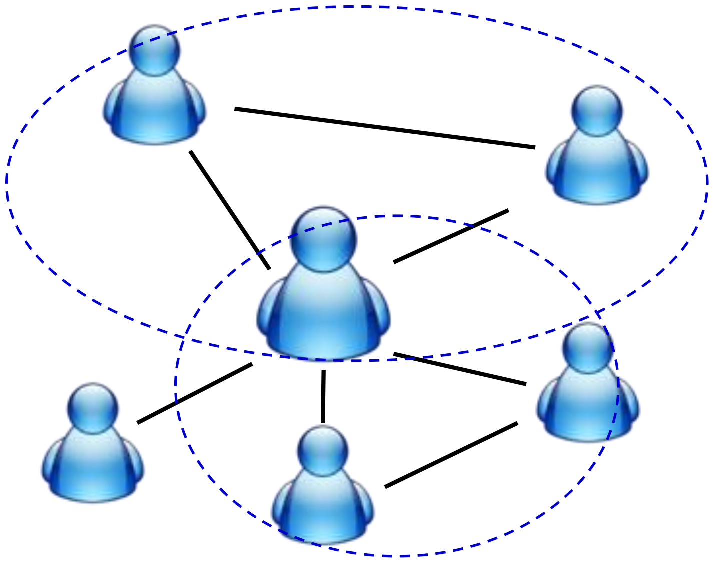
Web 3.0 = Social Web = Info. Space + Social Space



1. J. Scott. (1991, 2000, 2012). Social network analysis: A handbook.
2. D. Easley and J. Kleinberg. Networks, crowds, and markets: Reasoning about a highly connected world. Cambridge University Press, 2010.

Recent 5 years…

Web 4.0: Learning for the Web

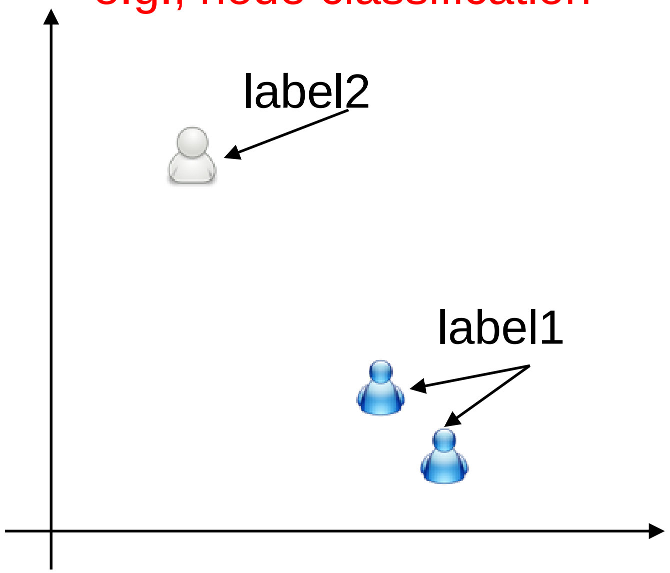


d-dimensional vector, *d*<<|*V*|

0.8	0.2	0.3	…	0.0	0.0

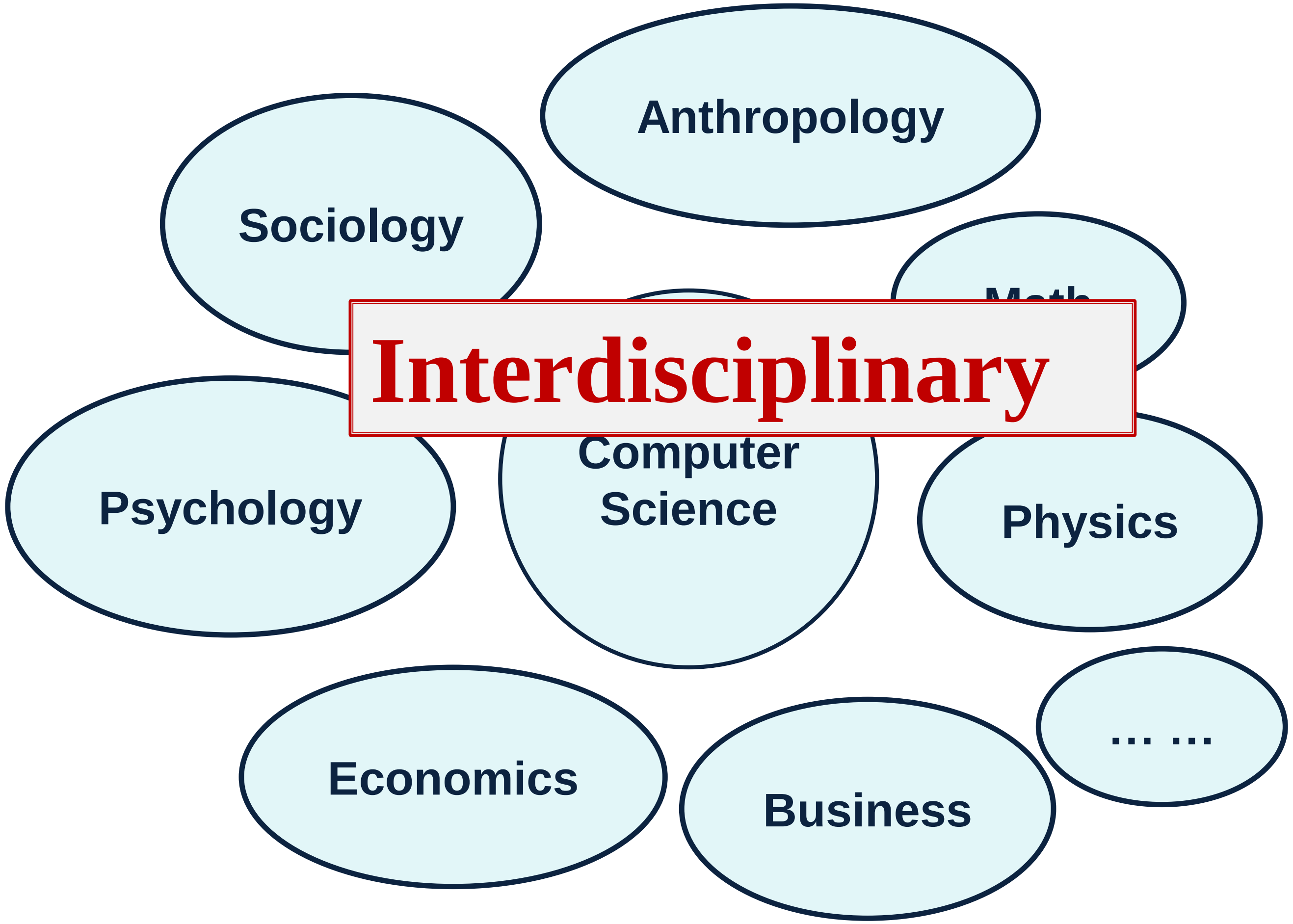
Users with the same label are located in the *d*-dimensional space closer than those with different labels

e.g., node classification



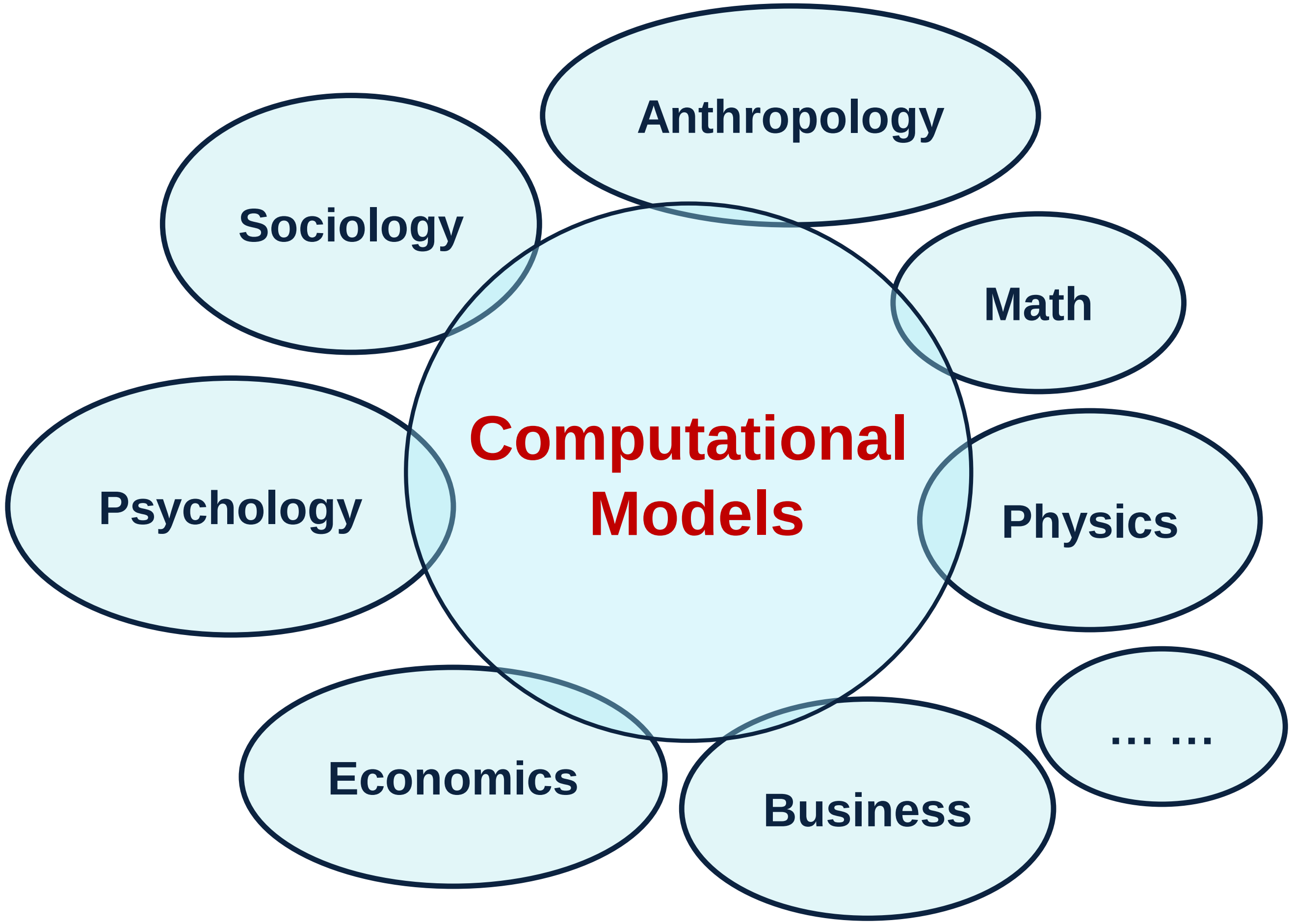
1. B. Perozzi, R. Al-Rfou, and S. Skiena. 2014. Deepwalk: Online learning of social representations. *KDD*, 701–710.

Social & Information Network Analysis



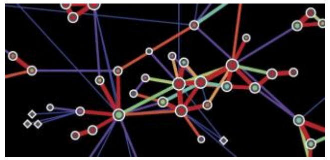
“A field is emerging that leverages the capacity to collect and analyze **data at a scale** that may reveal patterns of individual and group behaviors.”

Computational Models for Social & Information Network Analysis



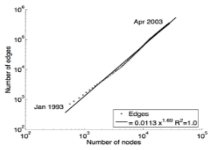
“A field is emerging that leverages the capacity to collect and analyze **data at a scale** that may reveal patterns of individual and group behaviors.”

Research of Social & Information Network Analysis



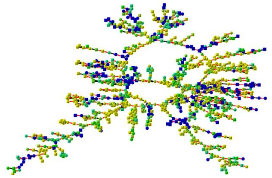
2015~2018

- Deep Learning for Networks
- High-Order Networks [Benson et al.]



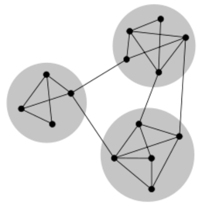
2010~2014

- Info. vs. Social Networks (Twitter) [Kwak et al.]
- Signed Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- Computational Social Science [Watts]
- Network Embedding [Perozzi et al.]



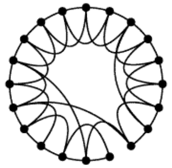
2005~2009

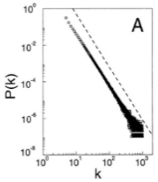
- Graph Evolution [Leskovec et al.]
- Three Deg. Of Influence [Christakis & Fowler]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Social Influence Analysis [Tang et al.]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]



2000~2004

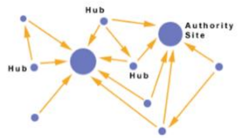
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]





1998/9

- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]

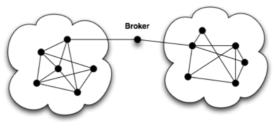


1997

- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]

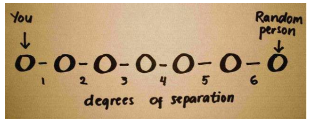
1992

- Structural Hole [Burt]
- Dunbar's Number [Dunbar]



1970s

- The Strength Of Weak Tie [Granovetter]

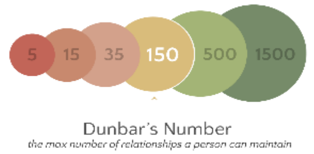


1960s

- Small Worlds [Migram]

1950s

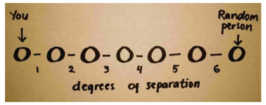
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]



1930s

- Sociogram [Moreno]

Research of Social & Information Network Analysis



- Small Worlds [Migram]

1960s

1950s

- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]

- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

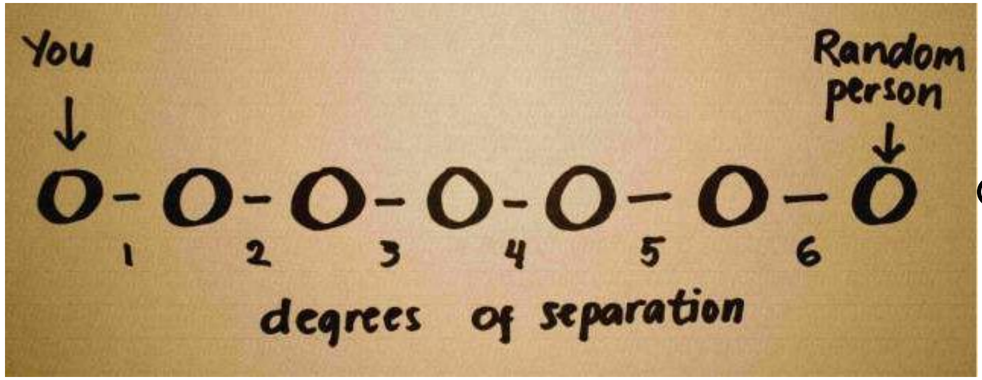
1930s

- Sociogram [Moreno]

1967: Six Degrees of Separation

- "Given two individuals selected randomly from the population, what is the probability that the minimum number of intermediaries required to link them is 0, 1, 2, …, *k*?"
- Milgram "selected 296 volunteers and to distribute a mail to a stockholder living in Boston."
- "The average number of intermediaries in the mailing chains was 5.2."

3.5 in 2016!



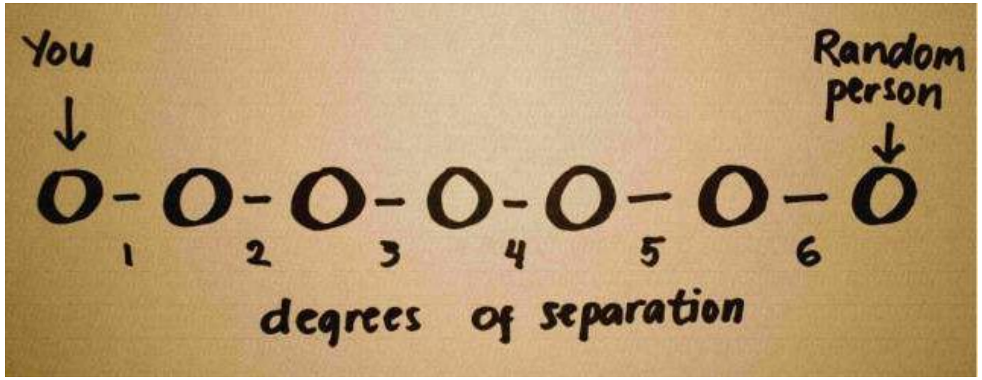
Q: What is the number (degree of separation) on Facebook?

1. Stanley Milgram. The small world problem. Psychology Today, 2(1):60–67, 1967. Cited by 8500+ (as of Aug 2018)

1967: Six Degrees of Separation

- "Given two individuals selected randomly from the population, what is the probability that the minimum number of intermediaries required to link them is 0, 1, 2, …, *k*?"
- Milgram "selected 296 volunteers and to distribute a mail to a stockholder living in Boston."
- "The average number of intermediaries in the mailing chains was 5.2."

3.5 in 2016!



Mark Zuckerberg

3.17 degrees of separation

1. Stanley Milgram. The small world problem. Psychology Today, 2(1):60–67, 1967. Cited by 8500+ (as of Aug 2018)

Research of Social & Information Network Analysis

2015~2018

2010~2014

2005~2009

2000~2004

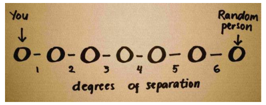
1998/9

1997

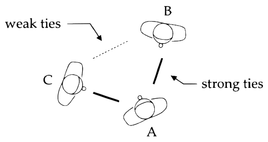
1992

1970s
- **The Strength Of Weak Tie [Granovetter]**

1960s
- Small Worlds [Migram]



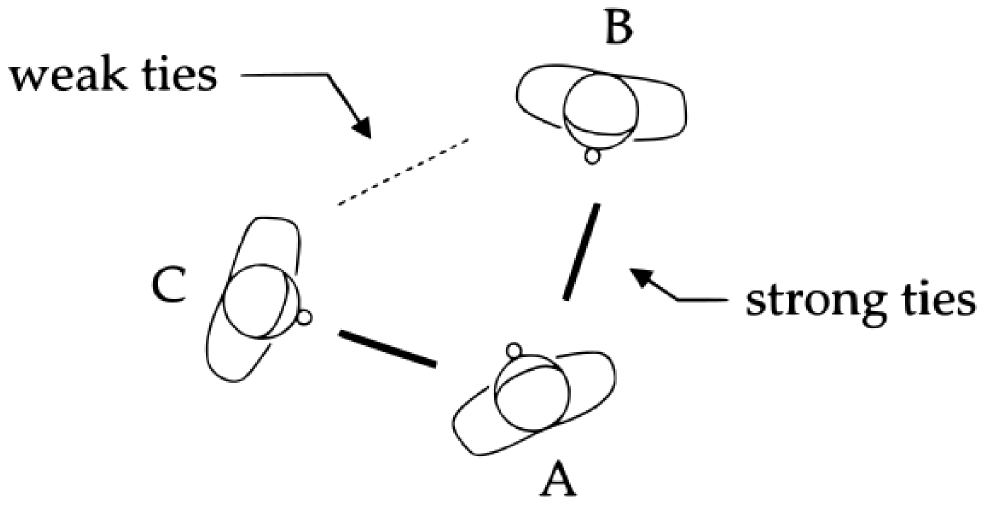
1950s
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]



1930s
- Sociogram [Moreno]

1973: Weak Tie

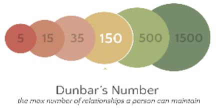
- The **weak tie hypothesis** argues, if A is linked to both B and C, then there is a greater-than-chance probability that B and C are linked to each other.
- Essentially form "A friend's friend is also my friend"
- Another important hypothesis based on weak tie is that information diffusion through weak ties rather than srong ties.



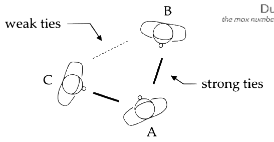
1. Granovetter, Mark S. The strength of weak ties. *American journal of sociology* 78.6 (1973): 1360-1380. Cited by 49000+ (as of Aug 2018)

Research of Social & Information Network Analysis

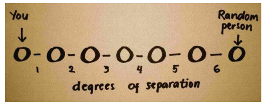
1992
- Structural Hole [Burt]
- **Dunbar's Number [Dunbar]**



1970s
- The Strength Of Weak Tie [Granovetter]



1960s
- Small Worlds [Migram]



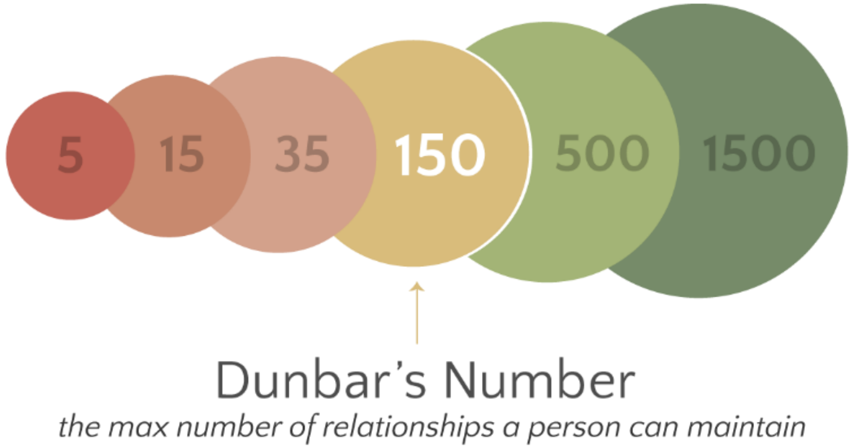
1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

1930s
- Sociogram [Moreno]

1992: Dunbar’s Number

- “**Dunbar‘s number** is a suggested cognitive limit to the number of people with whom one can maintain stable social relationships.”

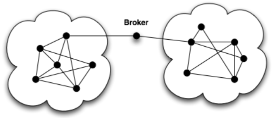
- Dunbar “proposed that humans can only comfortably maintain **150** stable relationships.”

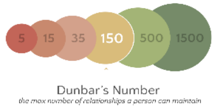


1. Robin I. M. Dunbar. Neocortex size as a constraint on group size in primates. *Journal of Human Evolution* 22 (6): 469–493. Cited by 2000+ (as of Aug 2018)

Research of Social & Information Network Analysis

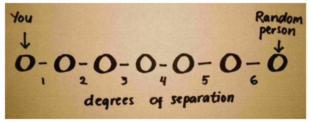
1992
- **Structural Hole [Burt]**
- Dunbar’s Number [Dunbar]



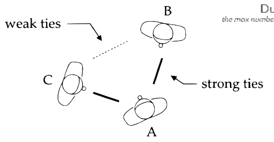


1970s
- The Strength Of Weak Tie [Granovetter]

1960s
- Small Worlds [Migram]



1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

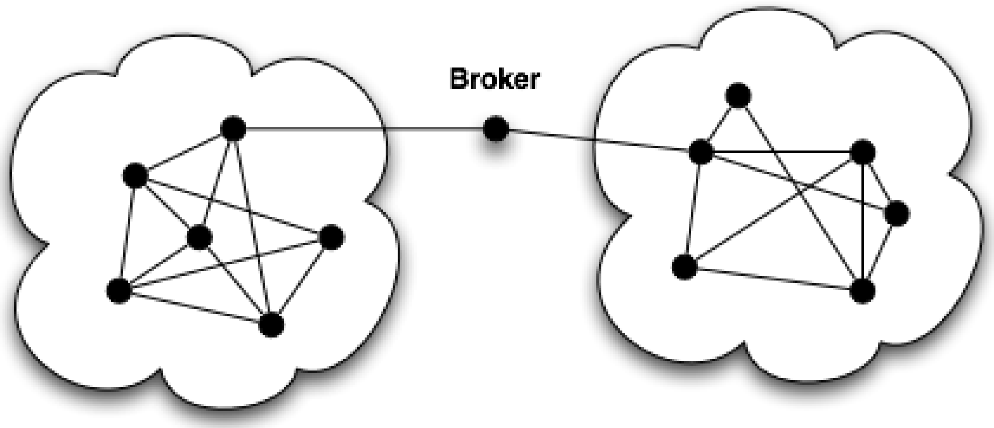


1930s
- Sociogram [Moreno]

1992/5: Structural Holes

- “The position of a bridge between distinct groups allows him or her to transfer valuable information from one group to another.”

- “The individual can combine all the ideas he or she receives from different sources and come up with the most innovative idea among all.”



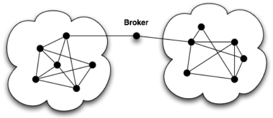
[1] Ronald S. Burt. *Structural Holes: The Social Structure of Competition*. Cambridge: Harvard University Press. 1995. Cited by 24000+ (as of Aug 2018)

Research of Social & Information Network Analysis

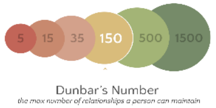
The 20th Century:

Sociology & Anthropology

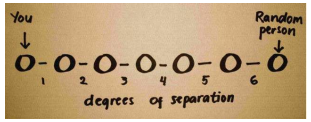
- 1992
  - Structural Hole [Burt]
  - Dunbar's Number [Dunbar]



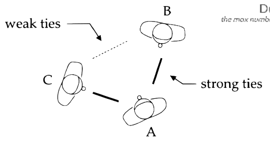
- 1970s
  - The Strength Of Weak Tie [Granovetter]



- 1960s
  - Small Worlds [Migram]



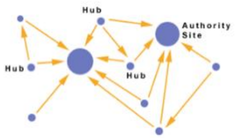
- 1950s
  - Homophily [Lazarsfeld & Merton]
  - Balance Theory [Heider et al.]
  - Random Graph [Erdos, Renyi, Gilbert]
  - Degree Sequence [Tuttle, Havel, Hakami]



- 1930s
  - Sociogram [Moreno]

Research of Social & Information Network Analysis

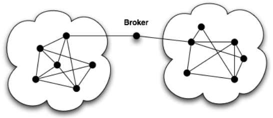
1997



- **HITS [Kleinberg]**
- **PageRank [Page & Brin]**
- Hyperlink Vector Voting [Li]

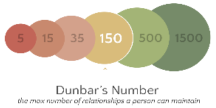
1992

- Structural Hole [Burt]
- Dunbar's Number [Dunbar]

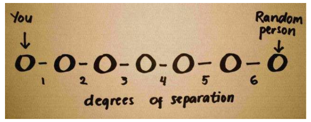


1970s

- The Strength Of Weak Tie [Granovetter]



1960s

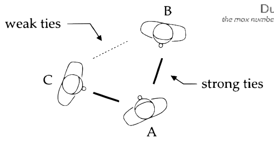


- Small Worlds [Migram]

1950s

- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]

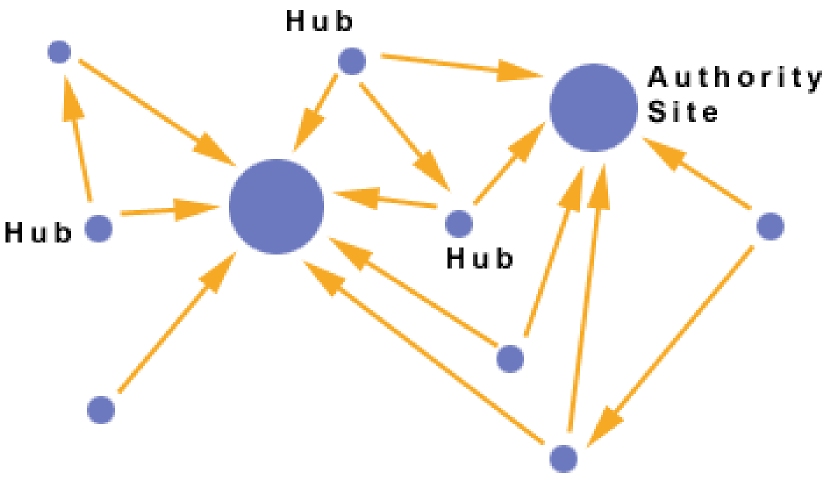
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

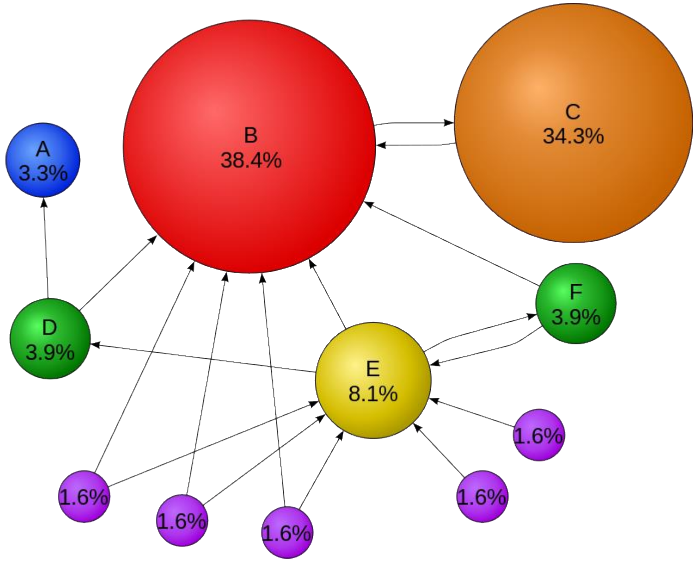


1930s

- Sociogram [Moreno]

1997-1998: HITS and PageRank





$$\text{auth}(p) = \sum_{i=1}^{n} \text{hub}(i)$$

$$\text{hub}(p) = \sum_{i=1}^{n} \text{auth}(i)$$

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

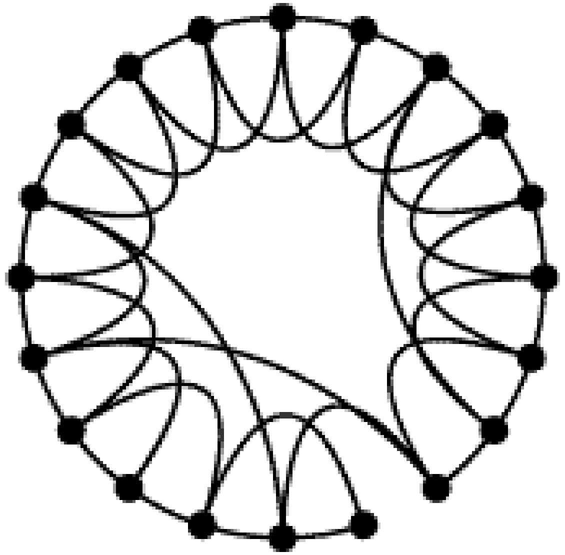
1. Jon M. Kleinberg. Authoritative sources in a hyperlinked environment. In ACM SODA, 1998. Also at IBM Research Report RJ 10076, May 1997. Cited by 12000+ (as of Aug 2018)
2. Sergey Brin, Lawrence Page. The anatomy of a large-scale hypertextual Web search engine. In WWW'07, Pages 107-117,1998. Cited by 17000+ (as of Aug 2018)

1998: Small World—Watts-Strogatz (WS) model

Comparing with the ER random graph model, WS model is a random graph generation model with Small-World properties.

Small-World Properties:

1. short average path lengths
2. high clustering coefficients



For each edge, rewire it with a probability

1. Duncan Watts, Steven Strogatz. *Collective dynamics of 'small-world' networks*. Nature 393 (6684): 440–442. Cited by 36000+ (as of Aug 2018)

1999: Scale Free—Barabási–Albert (BA) model

BA model is a random graph generation model using a preferential attachment mechanism: new nodes connect an existing node with the following probability:

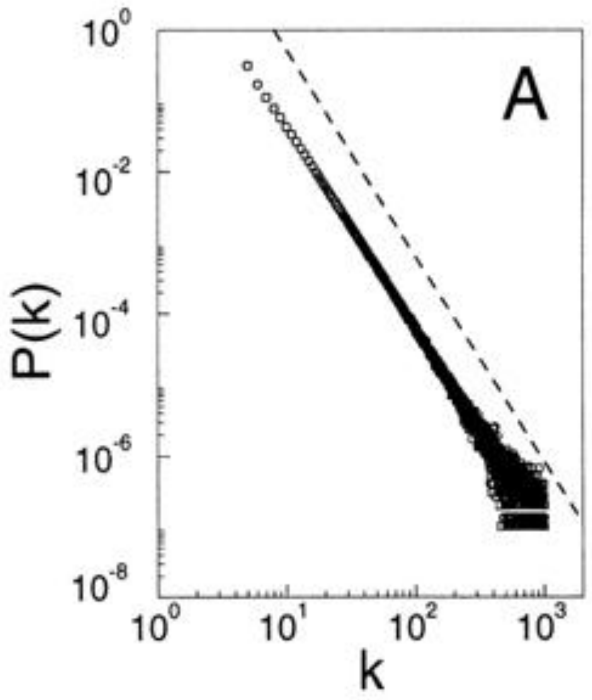
$$p_i = \frac{k_i}{\sum_j k_j}$$

- "A scale-free network is a network whose degree distribution follows a power law."

- The fraction $P(k)$ of nodes having k connections to other nodes:

$$\boldsymbol{P}(\boldsymbol{k}) \sim \boldsymbol{k}^{-\gamma}$$

where γ is a parameter whose value is typically in the range $2 < \gamma < 3$

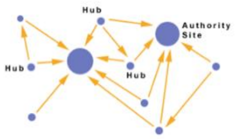


1. Albert-László Barabási, Réka Albert. Emergence of scaling in random networks, Science, 286:509–512, 1999. Cited by 31000+ (as of Aug 2018)
2. Michalis Faloutsos, Petros Faloutsos, Christos Faloutsos. On power-law relationships of the internet topology. In ACM SIGCOMM 1999. Cited by 6400+ (as of Aug 2018)

Research of Social & Information Network Analysis

The Late 20th Century:

CS & Physics



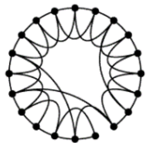
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]

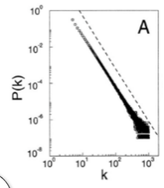
The 20th Century:

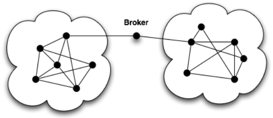
Sociology & Anthropology

1998/9
1997
1992
1970s
1960s
1950s
1930s

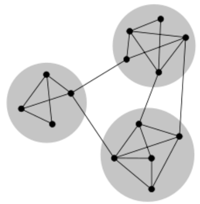
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]



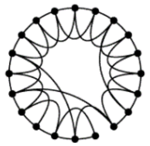


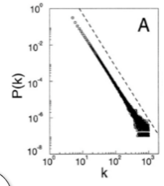


Research of Social & Information Network Analysis

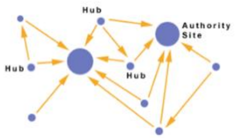


2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- **Community Detection [Girvan & Newman]**
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

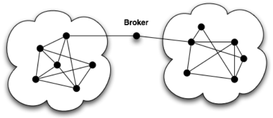




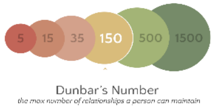
1998/9
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]



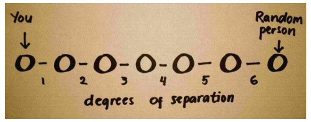
1997
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]



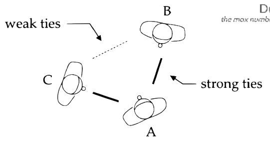
1992
- Structural Hole [Burt]
- Dunbar's Number [Dunbar]



1970s
- The Strength Of Weak Tie [Granovetter]



1960s
- Small Worlds [Migram]

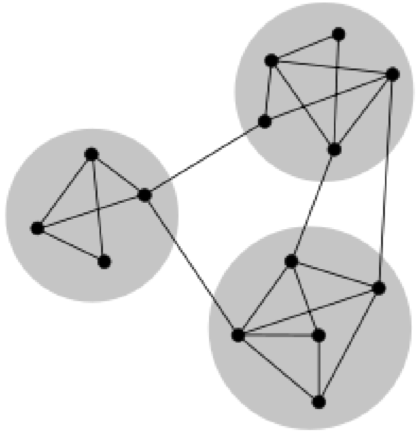


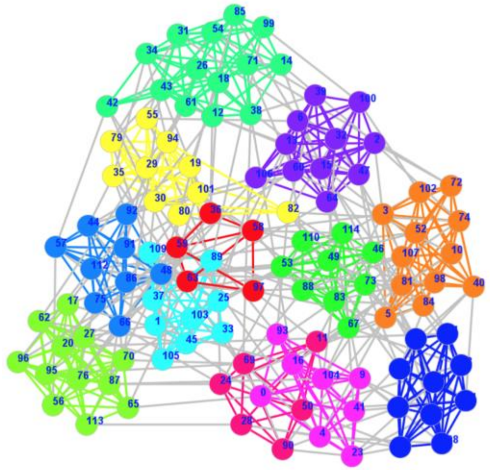
1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

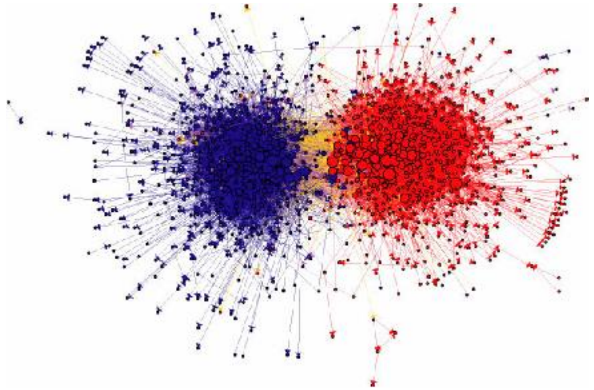
1930s
- Sociogram [Moreno]

2002: Community Detection

- “The property of community structure, in which network nodes are joined together in tightly knit groups, between which there are only looser connections.”



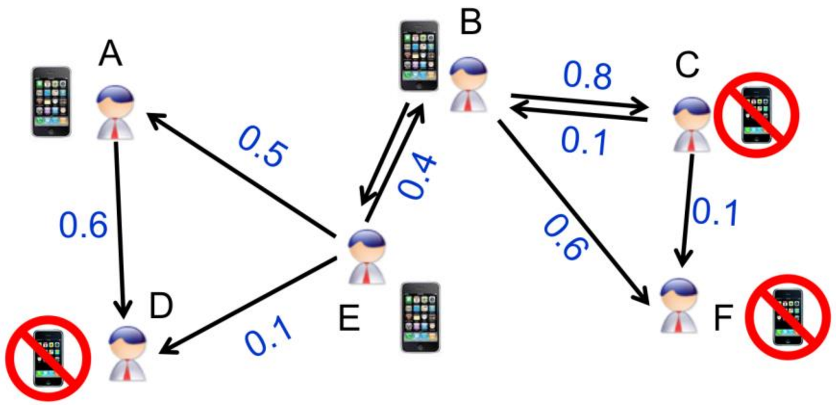




[1] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *PNAS 99(12):7821-7826, 2002*. Cited by 11000+ (as of Aug 2018)

2003: Influence Maximization

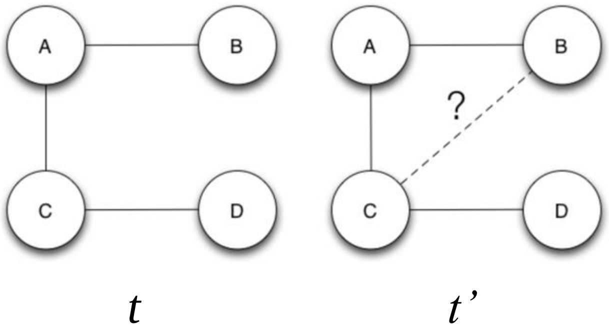
- Minimize marketing cost and more generally to maximize profit, e.g., to get a small number of influential users to adopt a new product, and subsequently trigger a large cascade of further adoptions.

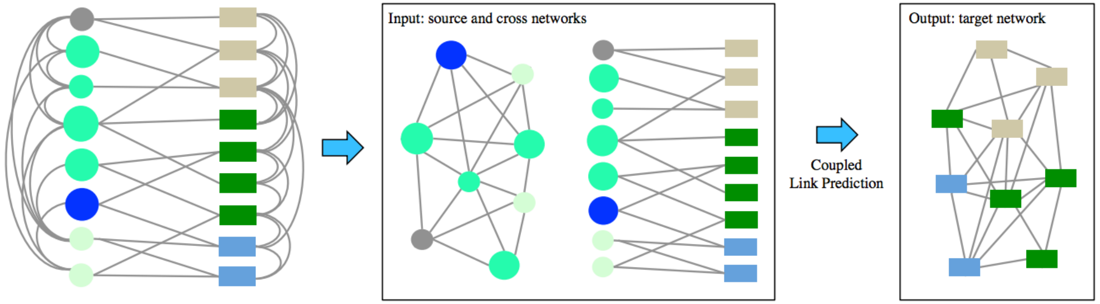


1. D. Kempe, J. Kleinberg, E. Tardos. *Maximizing the Spread of Influence through a Social Network*. In ACM KDD 2003. **Best Research Paper Award & Test of Time Paper Award**. Cited by 5600+ (as of Aug 2018)

2003: Link Prediction

- “Given a snapshot of a social network at time *t*, we seek to accurately predict the edges that will be added to the network during the interval from time *t* to a given future time *t’* .”

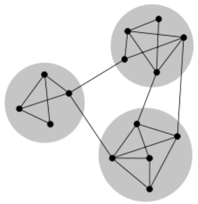




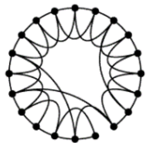
1. D. Liben-Nowell, J. Kleinberg. The Link Prediction Problem for Social Networks. In ACM CIKM, 2003. Cited by 3900+ (as of Aug 2018)

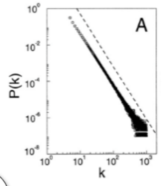
Research of Social & Information Network Analysis

2005~2009
- **Graph Evolution [Leskovec et al.]**
- Three Deg. Of Influence [Christakis & Fowler]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Social Influence Analysis [Tang et al.]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]

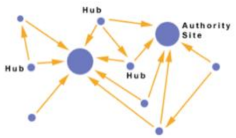


2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

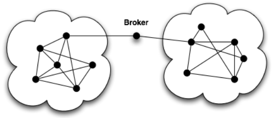


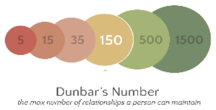


1998/9
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]

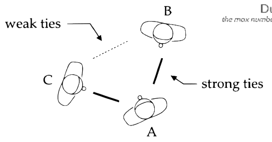


1997
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]

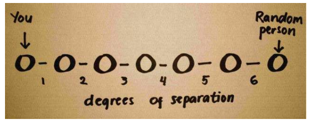




1992
- Structural Hole [Burt]
- Dunbar's Number [Dunbar]



1970s
- The Strength Of Weak Tie [Granovetter]



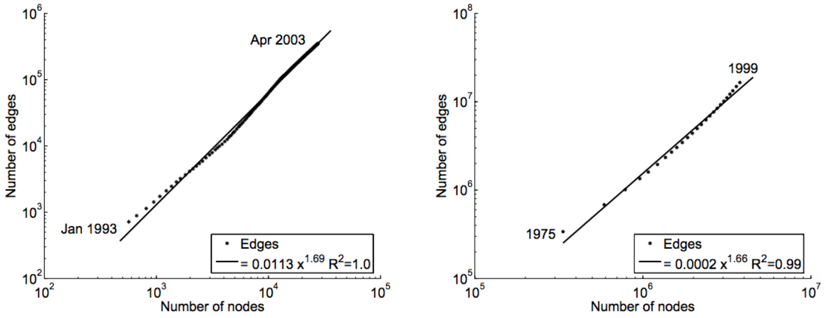
1960s
- Small Worlds [Migram]

1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

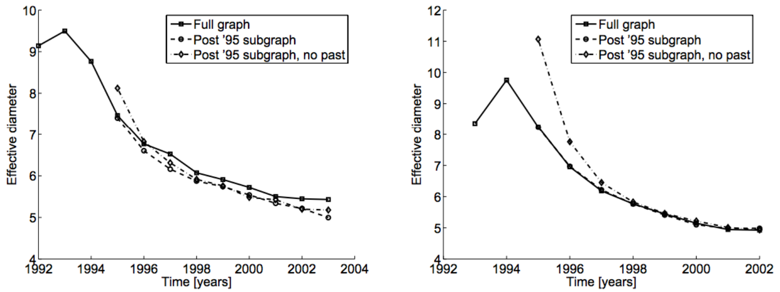
1930s
- Sociogram [Moreno]

2005: Network Evolution

- “Most of graphs densify over time, with the number of edges growing superlinearly in the number of nodes.”
- “The average distance between nodes often shrinks over time.”



Densification

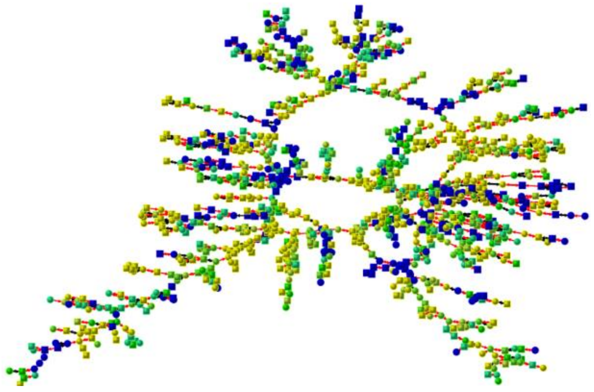


Shrinking Diameters

1. J. Leskovec, J. Kleinberg, C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. In KDD, 2005. **Best Research Paper Award & Test of Time Award**. Cited by 2000+ (as of Aug 2018)

2007: Diffusion and Influence

"If the husband became obese, the likelihood that his wife would become obese increased by **37%**."
—by tracking 10,000+ people for 32 years



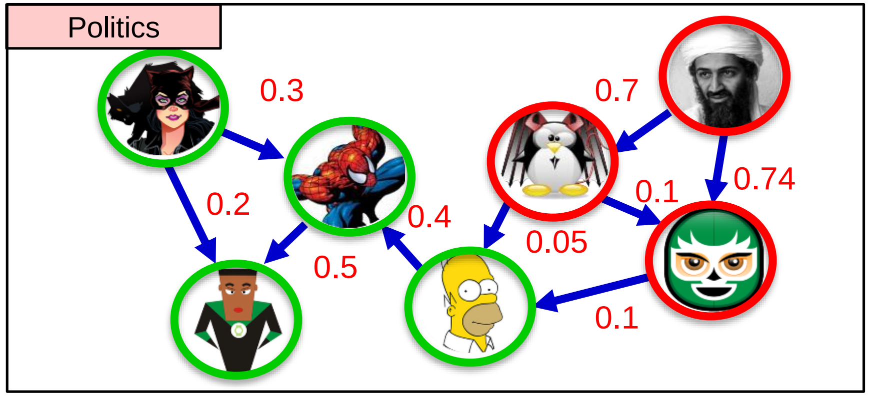
1. Nicholas Christakis, James Fowler. The Spread of Obesity in a Large Social Network Over 32 Years. The New England Journal of Medicine 357 (4): 370–379, 2007. Cited by 4600+ (as of May 2016)

2009: Social Influence Analysis

- "Social influence is a prevalent, complex and subtle force that governs the dynamics of all social networks."

- How to quantify the *social influences* from different topics?

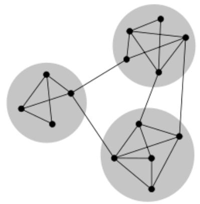
—Topical Affinity Propagation (TAP)



[1] Jie Tang, Jimeng Sun, Chi Wang, and Zi Yang. Social Influence Analysis in Large-scale Networks. In KDD'09, pages 807-816. **Top cited papers in KDD'09**. Cited by 700+ (as of Aug 2018)

Research of Social & Information Network Analysis

2005~2009
2000~2004
1998/9
1997
1992
1970s
1960s
1950s
1930s



The 1st decade of the 21st Century:

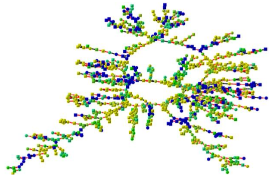
More Computer & Data Scientists

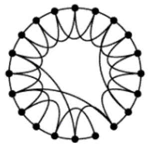
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

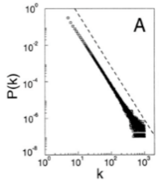
The 20th Century:

Sociology & Anthropology

- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social Influence Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]



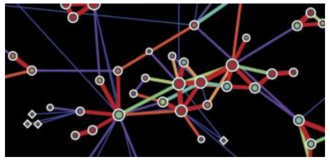


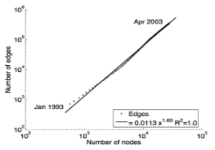


The Late 20th Century:

CS & Physics & Math

Research of Social & Information Network Analysis



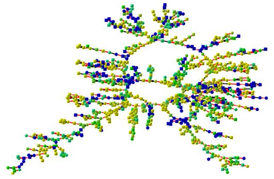


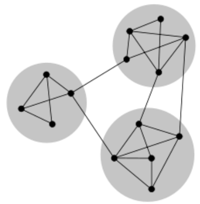
- Info. vs. Social Networks (Twitter) [Kwak et al.]
- Signed Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- **Computational Social Science [Giles]**
- Network Embedding [Perozzi et al.]

2010~2014

2005~2009

- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social Influence Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- **Computational Social Science [Lazer et al.]**



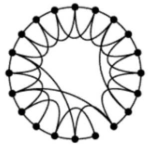


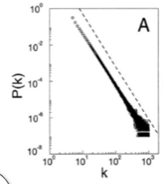
2000~2004

- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

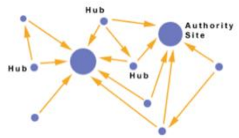
1998/9

- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]





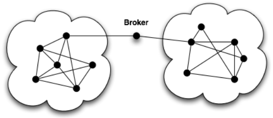
1997



- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]

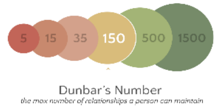
1992

- Structural Hole [Burt]
- Dunbar's Number [Dunbar]

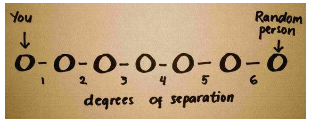


1970s

- The Strength Of Weak Tie [Granovetter]



1960s

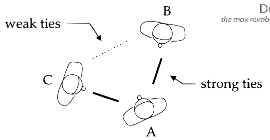


- Small Worlds [Migram]

1950s

- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]

- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]



1930s

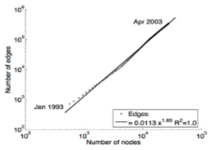
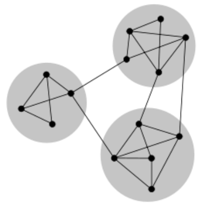
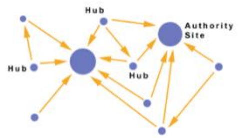
- Sociogram [Moreno]

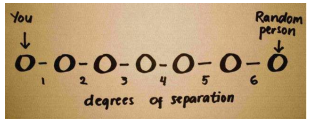
“A field is emerging that leverages the capacity to collect and analyze **data at a scale** that may reveal patterns of individual and group behaviors.”

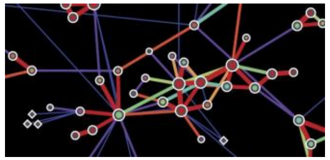
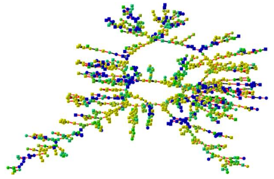
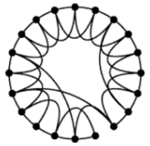
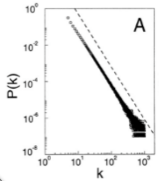
—*David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-Laszlo Barabasi,* et al. from *Sociology, Computer Science, Physics, Business, Government*, etc. at *Harvard, MIT, Northeastern, Northwestern, Columbia, Cornell, etc.*

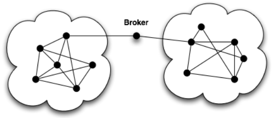
1. **David Lazer et al. Computational Social Science. *Science* 2009.**
2. James Giles. Computational Social Science: Making the Links. **Nature** 2012.

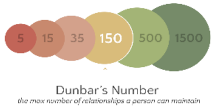
Research of Social & Information Network Analysis

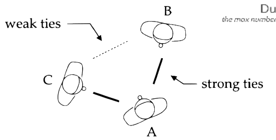












2015~2018
- **Deep Learning for Networks**
- High-Order Networks [Benson et al.]

2010~2014
- Info. vs. Social Networks (Twitter) [Kwak et al.]
- Signed Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- Computational Social Science [Watts]
- **Network Embedding [Perozzi et al.]**

2005~2009
- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social Influence Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]

2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

1998/9
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]

1997
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]

1992
- Structural Hole [Burt]
- Dunbar's Number [Dunbar]

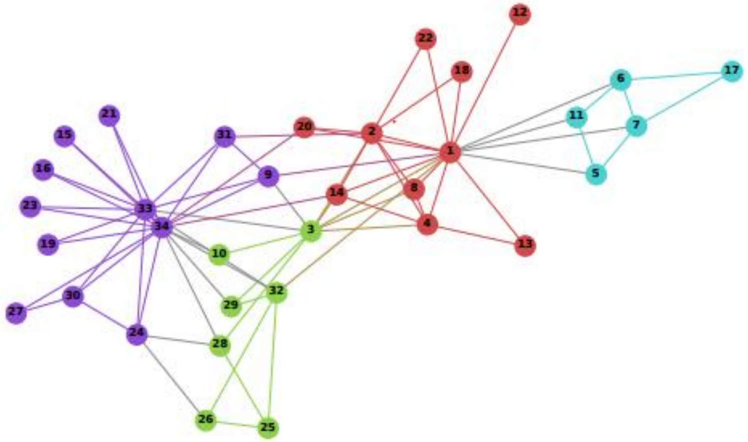
1970s
- The Strength Of Weak Tie [Granovetter]

1960s
- Small Worlds [Migram]

1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

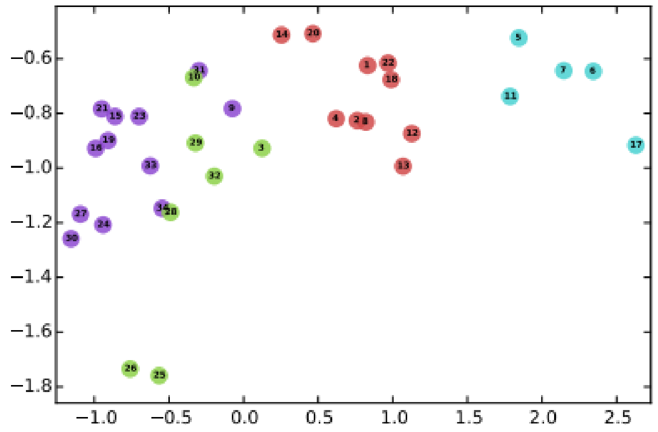
1930s
- Sociogram [Moreno]

Network Representation Learning



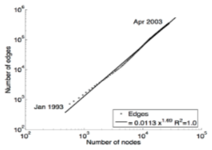
⇨

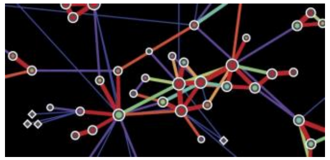
d-dimensional vector, *d*<<|*V*|



1. Lei Tang and Huan Liu. Relational learning via latent social dimensions. In *KDD 2009.*
2. Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. DeepWalk: Online learning of social representations. In *KDD 2014*. **The most cited paper in KDD’14**. (as of Aug 2018)

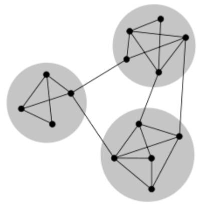
Research of Social & Information Network Analysis

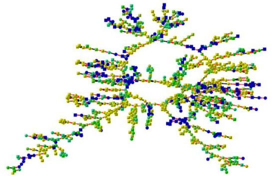




2015~2018
- **Deep Learning for Networks**
- High-Order Networks [Benson et al.]

2010~2014
- Info. vs. Social Networks (Twitter) [Kwak et al.]
- Signed Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- Computational Social Science [Watts]
- Network Embedding [Perozzi et al.]

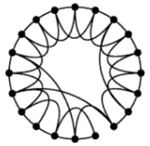


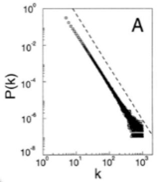


2005~2009
- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social Influence Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Horvitz]
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]

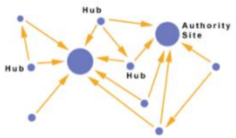
2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

1998/9
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]

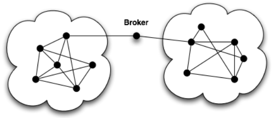




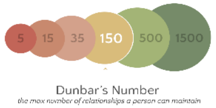
1997
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]



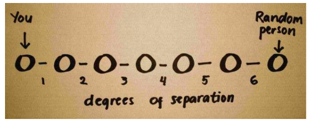
1992
- Structural Hole [Burt]
- Dunbar's Number [Dunbar]



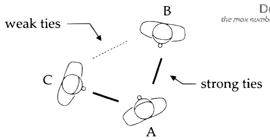
1970s
- The Strength Of Weak Tie [Granovetter]



1960s
- Small Worlds [Migram]

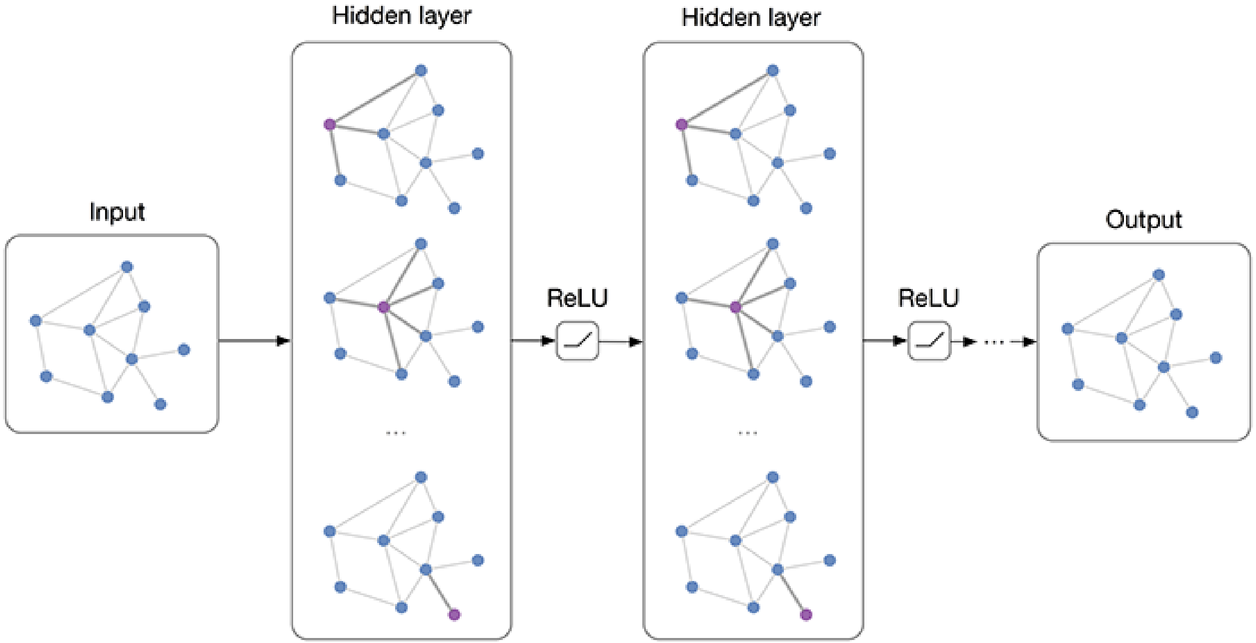


1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]



1930s
- Sociogram [Moreno]

Deep networks for networks

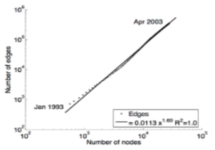


1. Kipf & Welling, Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR 2017.*
2. Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Y Bengio. Graph Attention Networks. In *ICLR 2018.*

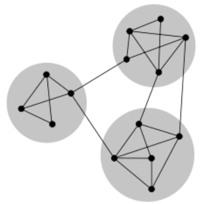
Research of Social & Information Network Analysis

2015~2018
- Deep Learning for Networks
- High-Order Networks [Benson et al.]

Recent Trend:
Deep Learning for Networks



2010~2014
- Info. vs. Social Networks (Twitter) [Kwak et al.]
- Signed Networks [Leskovec et al.]
- Semantic Social Networks [Tang et al.]
- Four Deg. Of Separation [Backstrom et al.]
- Structural Diversity [Ugander et al.]
- Computational Social Science [Watts]
- Network Embedding [Perozzi et al.]



2005~2009
- Graph Evolution [Leskovec et al.]
- 3 Deg. Of Influence [Christakis & Fowler]
- Social Influence Analysis [Tang et al.]
- Six Deg. Of Separation [Leskovec & Hor
- Network Heterogeneity [Sun & Han]
- Network Embedding [Tang & Liu]
- Computer Social Science [Lazer et al.]

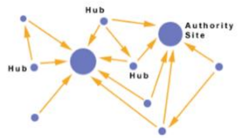
The 1st decade of the 21st Century:
More Computer & Data Scientists

2000~2004
- Influence Max'n [Domingos & Kempe et al.]
- Community Detection [Girvan & Newman]
- Network Motifs [Milo et al.]
- Link Prediction [Liben-Nowell & Kleinberg]

1998/9
- Small Worlds [Watts & Strogatz]
- Scale Free [Barabasi & Albert]
- Power Law [Faloutsos × 3]

The Late 20th Century:
CS & Physics

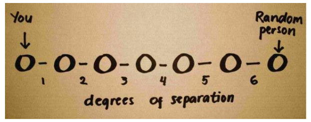
1997
- HITS [Kleinberg]
- PageRank [Page & Brin]
- Hyperlink Vector Voting [Li]



1992
- Structural Hole [Burt]
- Dunbar's Number [Dunbar]

1970s
- The Strength Of Weak Tie [Granovetter]

1960s
- Small Worlds [Migram]

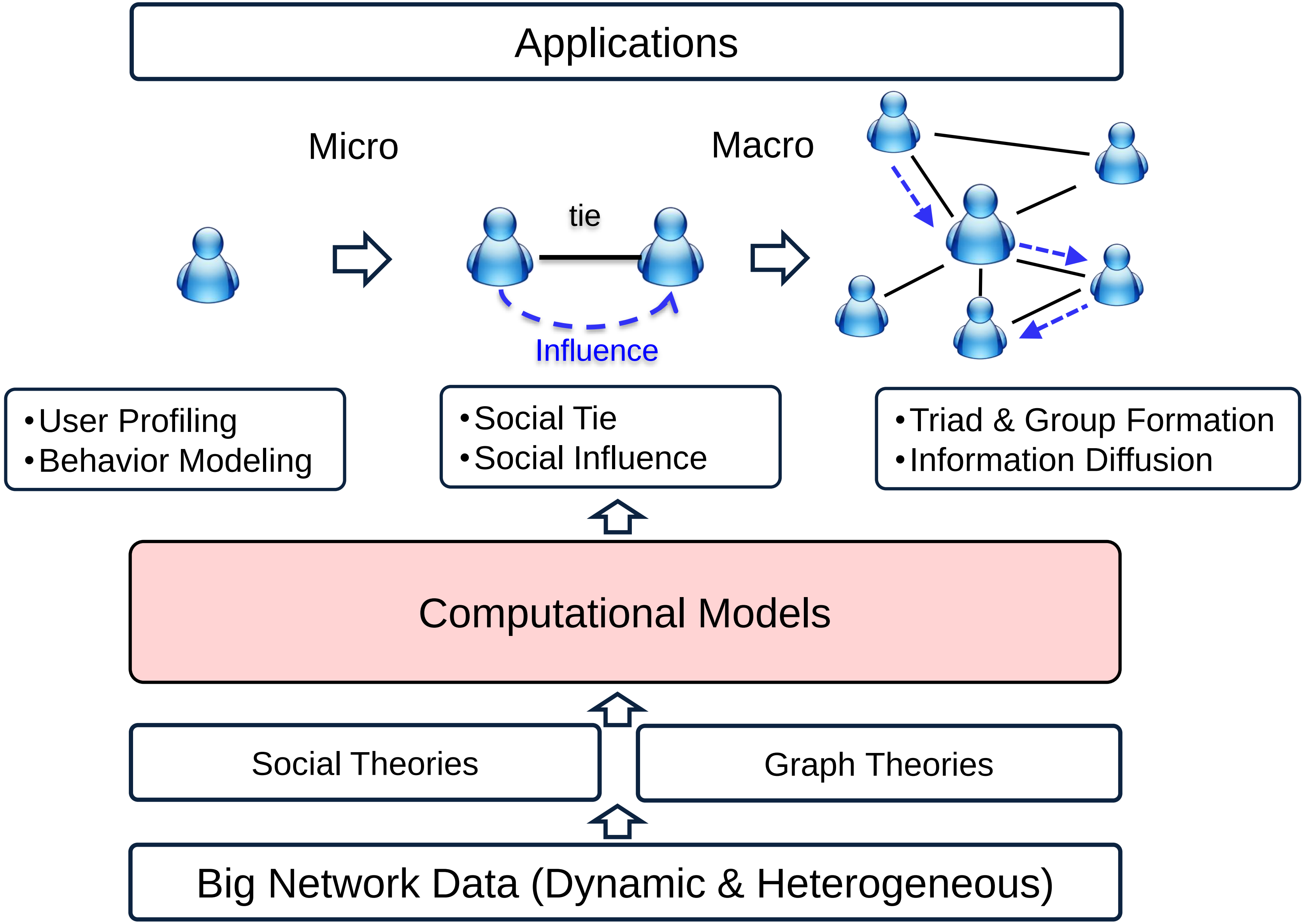


1950s
- Random Graph [Erdos, Renyi, Gilbert]
- Degree Sequence [Tuttle, Havel, Hakami]
- Homophily [Lazarsfeld & Merton]
- Balance Theory [Heider et al.]

1930s
- Sociogram [Moreno]

The 20th Century:
Sociology & Anthropology

Social & Information Network Analysis



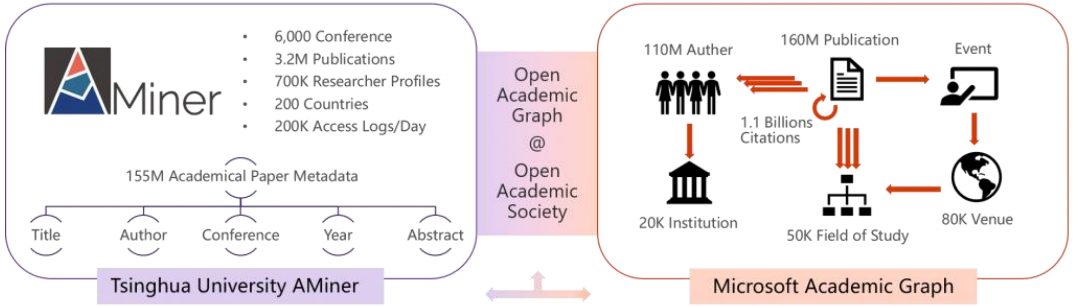
Representation Learning on Networks

- **The first part:**
  - Conventional network analysis
    - Node classification
    - Social tie & link prediction
  - Network embeddings
    - Embedding models
    - Theoretical understanding
    - Large-scale embedding
- **The second part:**
  - Graph neural networks
    - Graph convolution
    - Graph GAN
    - Dynamic Representation
    - Heterogeneous Representation
  - Large-scale applications
    - Knowledge graph linking
    - Recommendation in E-commerce
    - Online-to-offline recommendations
    - Social influence in gaming

Applications in this talk

OAG: Open Academic Graph

https://www.openacademic.ai/oag/



Data set	#Pairs/Venues	Date
Linking relations	29,841	2018.12
AMiner venues	69,397	2018.07
MAG venues	52,678	2018.11

Table 1: statistics of OAG venue data

Data set	#Pairs/Papers	Date
Linking relations	91,137,597	2018.12
AMiner papers	172,209,563	2019.01
MAG papers	208,915,369	2018.11

Table 2: statistics of OAG paper data

Data set	#Pairs/Authors	Date
Linking relations	1,717,680	2019.01
AMiner authors	113,171,945	2018.07
MAG authors	253,144,301	2018.11

Open Academic Graph

Open Academic Graph (OAG) is a large knowledge graph unifying two billion-scale academic graphs: Microsoft Academic Graph (MAG) and AMiner. In mid 2017, we published OAG v1, which contains 166,192,182 papers from MAG and 154,771,162 papers from AMiner (see below) and generated 64,639,608 linking (matching) relations between the two graphs. This time, in OAG v2, author, venue and newer publication data and the corresponding matchings are available.

Overview of OAG v2

The statistics of OAG v2 is listed as the three tables below. The two large graphs are both evolving and we take MAG November 2018 snapshot and AMiner July 2018 or January 2019 snapshot for this version.

1. Zhang, et al. OAG: Toward Linking Large-scale Heterogeneous Entity Graphs. In KDD'19.

Datasets for Social Network Analysis

Datasets for Social Network Analysis

https://www.aminer.cn/data-sna

SN	Name	Node	Edge	Behavior/Content	Description
			Microblogging networks		
1	Twitter-Dynamic-Net	90908 users	443,399 time varying following relationships	99,696,204 tweets associated with 156,487 users	Dynamic twitter following network and tweets.
2	Twitter-Dynamic-Action	7514 users	304,275 time varying following relationships	730,568 tweets	Tweeting actions of users on a specific topic "Haiti Earthquake"
3	Twitter-Competitor	87,603 Twitter users		1,033,750 tweets covers 1393 companies	Twitter content related to companies
4	Twitter-Net-Tweet	41.7 million users	1.47 billion social relations	4,252 trending topics, 106 million tweets	The entire Twitter site in 2010
5	Weibo-Net-Tweet	1,776,950	308,489,739	300,000 original microblogs and 23,755,810 retweets	Sina weibo users, relationships, and their tweets and retweets.
			Patent data set from Patentminer.org		
6	Patent	2,334,093 inventors	11,504,051 coauthor relationships	4,179,629 patents and 584,380 companies	Co-patent and patent citation network
			Other online social networks		
7	Slashdot-large	93139 users	577025 friend/foe relationships	35065 news and 3505736 comments	Slashdot friend and foe network and news comment data
8	Slashdot-small	13,182 users	309,14 friends, 5,424 foes		Slashdot friend and foe network
9	Epinions-1	25,148 users	74,060 trust relationships, 31,001 distrust relationships		Epinioins trust/distrust network
10	Epinions-2	22,166 users	355,813 links between users	296,277 items, 27 categories, 922,267 ratings	Epinions trust/distruct network and user rating item data
11	Enron	151 users	133 manager-subordinate relationships, 132 colleague relationships		Email communication network
12	Flickr-large	2,037,538 users	219,098,660 friend relationships	655,917 groups, 1,262,978 images, 1,4913,164 user comments	Flick friend network and image comments
13	Flickr-medium	215,495 users	9,114,557 relationships		Flick friend network

Representation Learning on Networks

- **The first part:**
  - Conventional network analysis
    - Node classification
    - Social tie & link prediction
    - Influence & behavior modeling
  - Network embeddings
    - Embedding models
    - Theoretical understanding
    - Large-scale embedding
- **The second part:**
  - Graph neural networks
    - CNN → Graph convolution
    - GAN → Graph GAN
  - Large-scale applications
    - Knowledge graph linking
    - Recommendation in E-commerce
    - Online-to-offline recommendations
    - Social influence in gaming

Thank you！

Collaborators: John Hopcroft, Jon Kleinberg, Chenhao Tan (**Cornell**)

Jiawei Han (**UIUC**), Philip Yu (**UIC**)

Jian Pei (**SFU**), Hanghang Tong (**ASU**)

Tiancheng Lou (**Google&Baidu**), Jimeng Sun (**GIT**)

Wei Chen, Ming Zhou, Long Jiang, Chi Wang, Kuansan Wang (**Microsoft**)

Hongxia, Jingren Zhou, Chang Zhou (**Alibaba**)

Jiezhong Qiu, Jie Zhang, Fanjin Zhang, Qibin Chen, Yukuo Cen, et al. (**THU**)

Jie Tang, KEG, Tsinghua U, http://keg.cs.tsinghua.edu.cn/jietang

Download all data & Codes, http://arnetminer.org/data

http://arnetminer.org/data-sna