Representation Learning on Networks

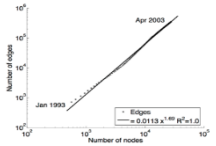
Algorithms, Theory, and Applications

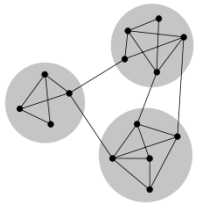
Jie Tang
Tsinghua University

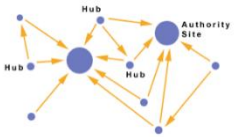
Yuxiao Dong
Microsoft Research, Redmond

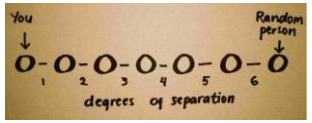
Microsoft Research

Research of Social & Information Network Analysis

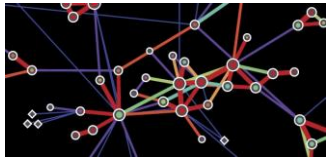


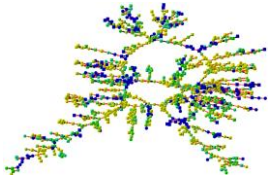


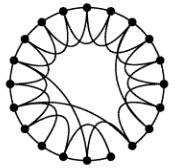


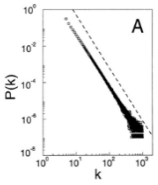


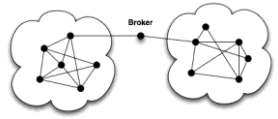
- 2015~2018
  - Deep Learning for Networks
  - High-Order Networks [Benson et al.]
- 2010~2014
  - Info. vs. Social Networks (Twitter) [Kwak et al.]
  - Signed Networks [Leskovec et al.]
  - Semantic Social Networks [Tang et al.]
  - Four Deg. Of Separation [Backstrom et al.]
  - Structural Diversity [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]

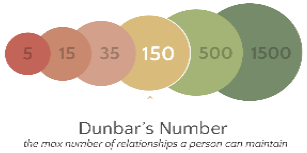




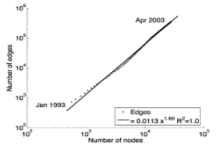


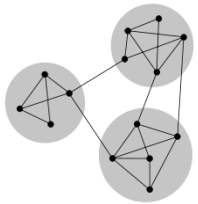


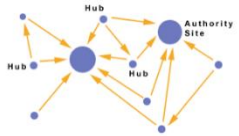


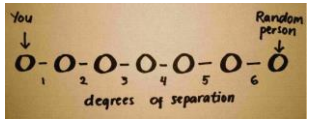


Research of Social & Information Network Analysis









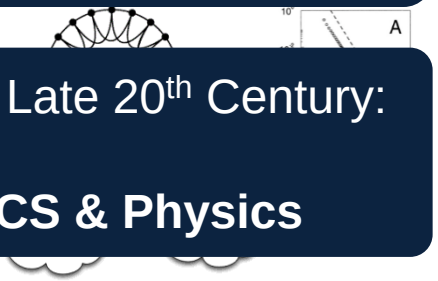
Recent Trend: Deep Learning for Networks

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

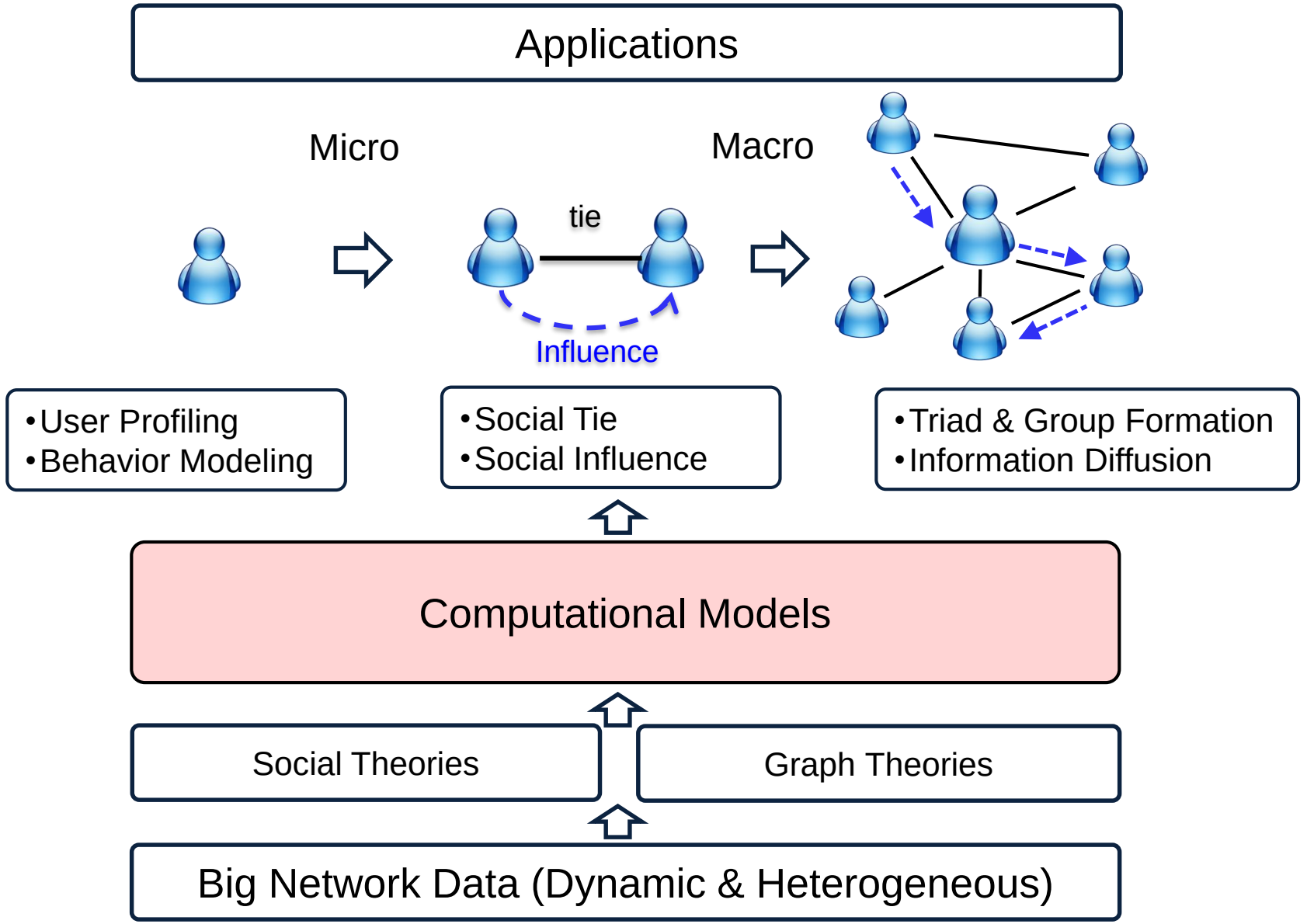
The Late 20th Century: CS & Physics

The 20th Century: Sociology & Anthropology

- **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 & Hor
  - 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**
  - Homophily [Lazarsfeld & Merton]
  - Balance Theory [Heider et al.]
- **1930s**
  - Sociogram [Moreno]



Social & Information Network Analysis



Representation Learning on Networks

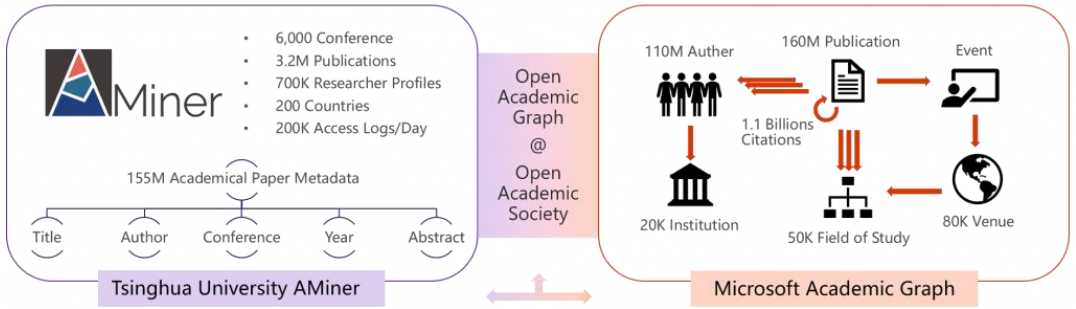
- **The first part:**
  - Conventional network analysis
    - Node classification
    - Social tie & link prediction
  - Network embeddings
    - Embedding models
    - Theoretical understanding
    - Large-scale embedding

- **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

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,

Download all data & Codes,

http://keg.cs.tsinghua.edu.cn/jietang

http://arnetminer.org/data

http://arnetminer.org/data-sna