EMERGING NETWORK SCIENCE APPLICATIONS IN FINANCE & EDUCATION
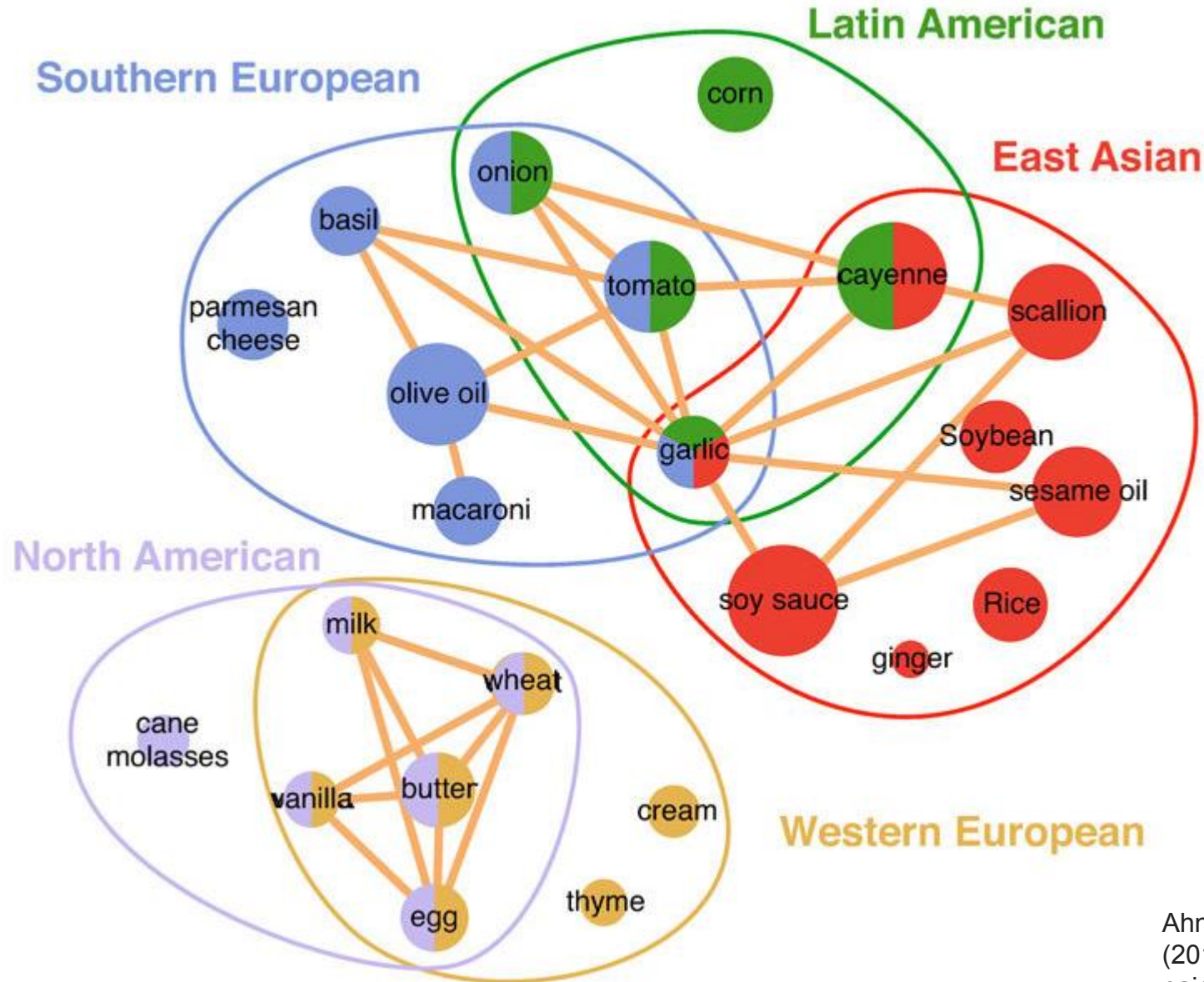
Dr. Prasanta BHATTACHARYA [1,2]

[1]Social & Cognitive Computing | IHPC
[2]Analytics & Operations | NUS Business School
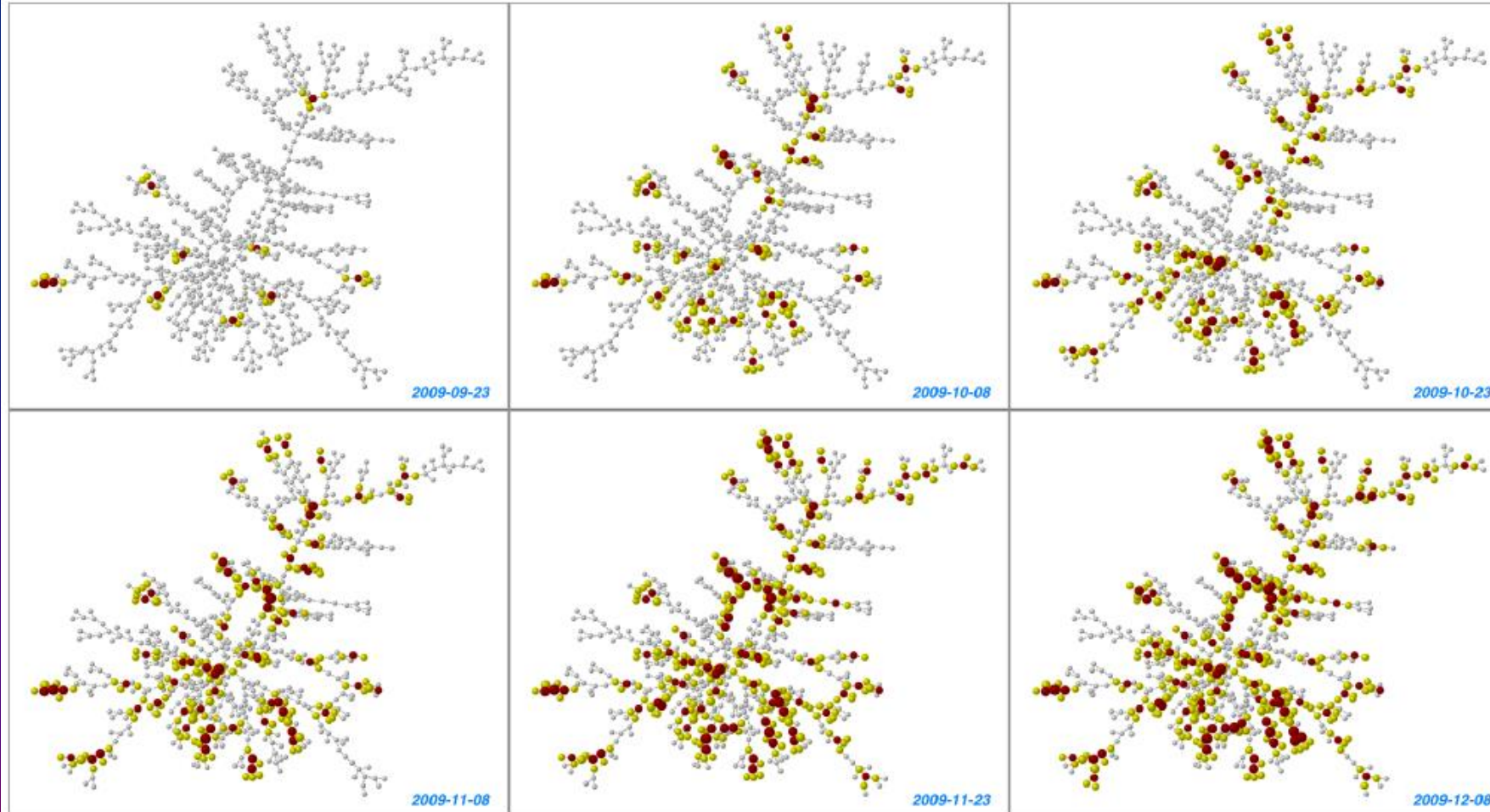
Sept 23, 2020

Co-occurrence in recipes

Networks help us understand phenomena in new light

Ahn, Y. Y., Ahnert, S. E., Bagrow, J. P., & Barabási, A. L. (2011). Flavor network and the principles of food pairing. *Scientific reports*, *1*, 196.

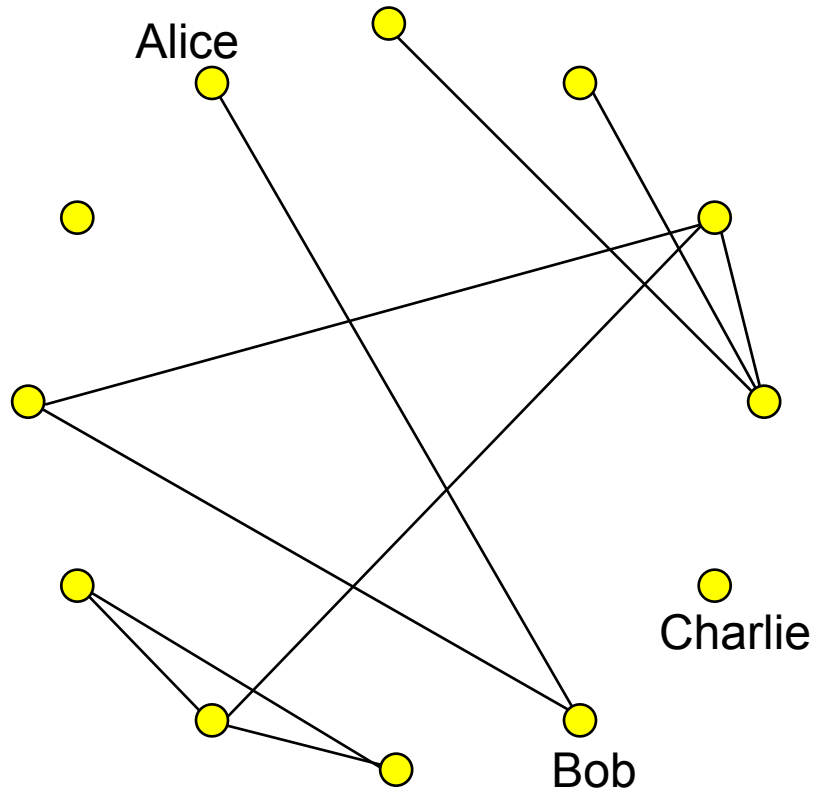# During public health crises, network science plays a key role

Peak incidence of epidemic among friends occurred ~14 days in advance of a randomly selected groups

https://medium.com/@nicholas.christakis/how-social-networks-can-detect-epidemics-in-advance-f3de6ada97bb

Networks 101

- Individual actors are represented by nodes
- Relationships are represented as edges

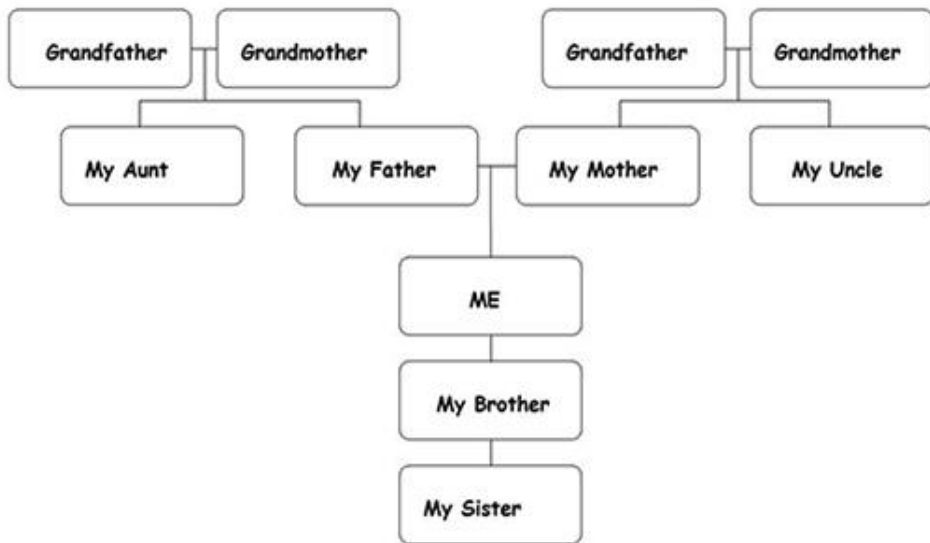(relationships maybe friendship, kinship, co-authors, similarity etc.)

Prediction: Who will Alice befriend next?

Explanation: Why/how did Alice befriend Bob?

Alice

Charlie

Bob

Network sizes vary substantially depending on context

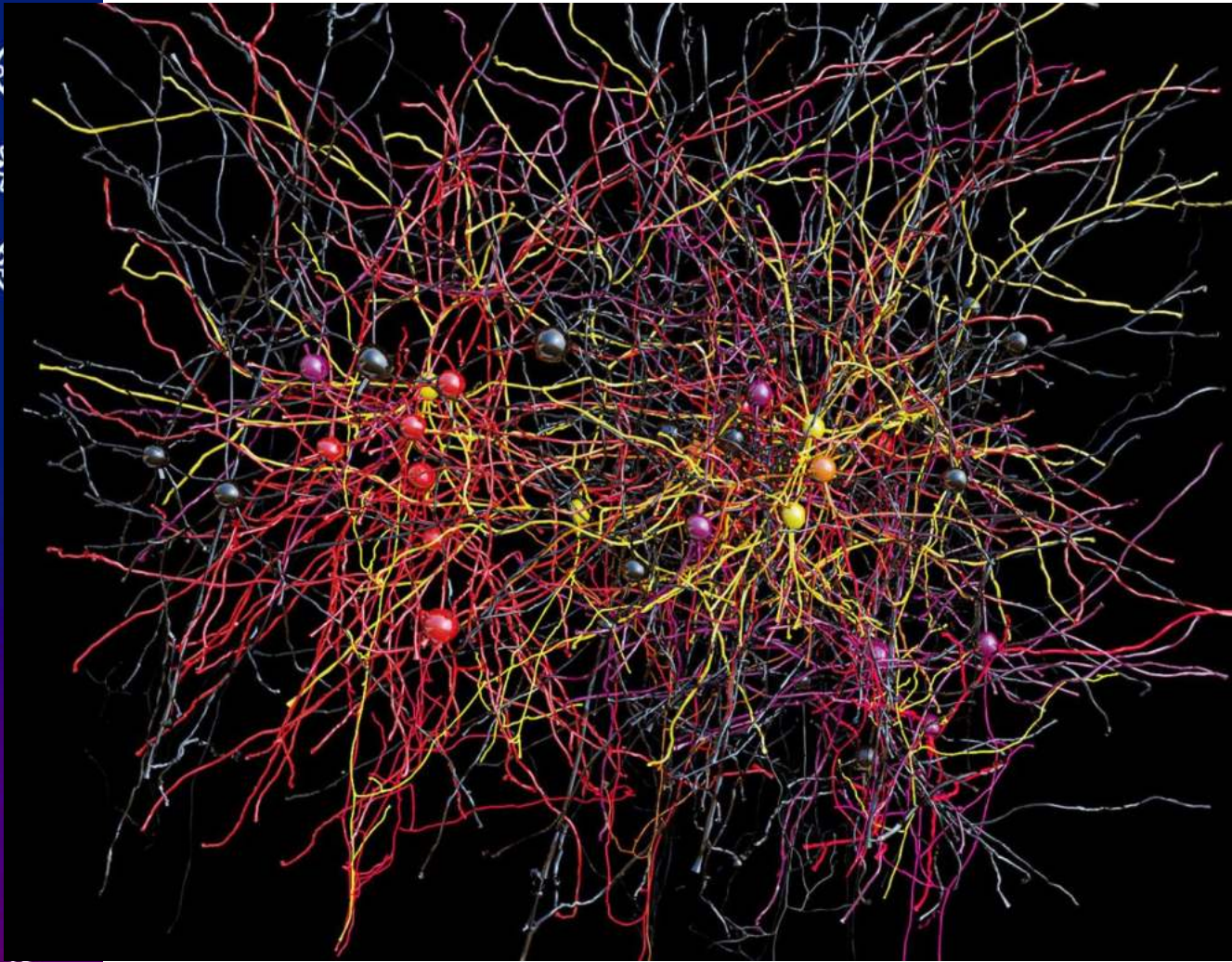**Current family network**

N ~ 20-30

Facebook graph

N ~ 2 billion

http://www.open.edu/openlearncreate/mod/oucontent/view.php?id=80895&section=1.5

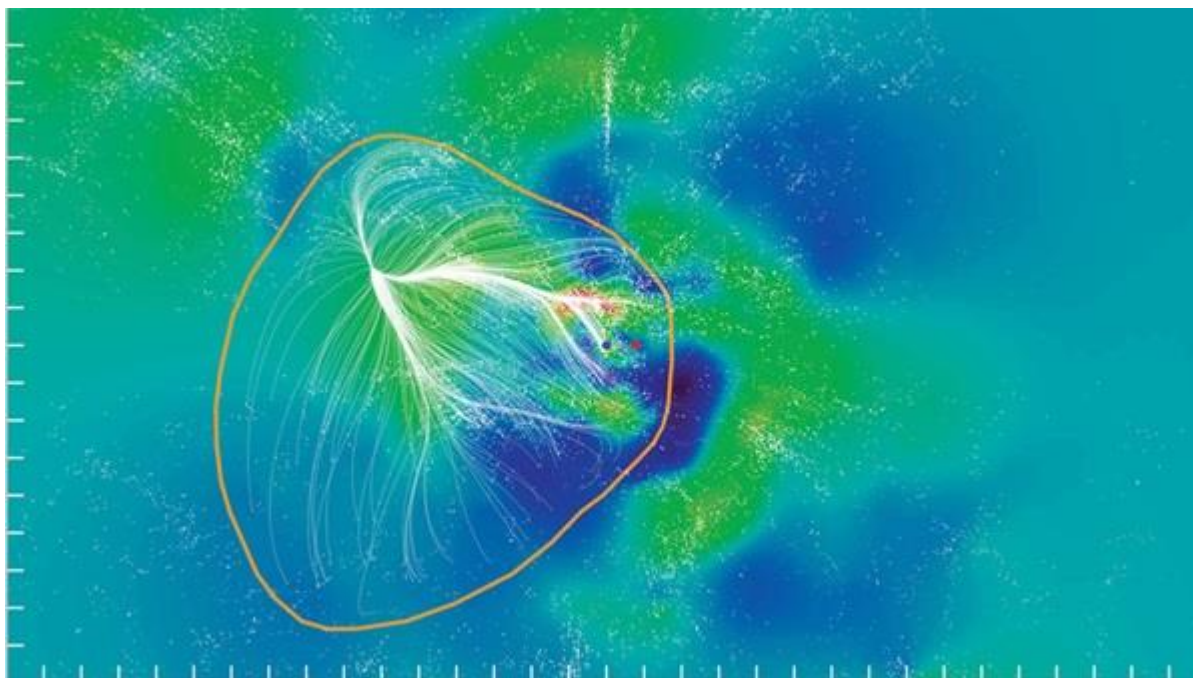https://www.facebook.com/notes/facebook-engineering/visualizing-friendships/469716398919

- Landmark study connecting brain network to specific function
- Largest fully mapped neural network
- 1,278 connected neurons

https://www.wired.com/2016/03/took-neuroscientists-ten-years-map-tiny-slice-brain/

450 X 450 X 150 micro m

Newly identified galactic supercluster is home to the Milky Way

This discovery clarifies the boundaries of our galactic neighborhood and establishes previously unrecognized linkages among various galaxy clusters in the local universe.

By NRAO, Charlottesville, VA | Published: Thursday, September 4, 2014

http://astronomy.com/news/videos/2014/09/newly-identified-galactic-supercluster-is-home-to-the-milky-way
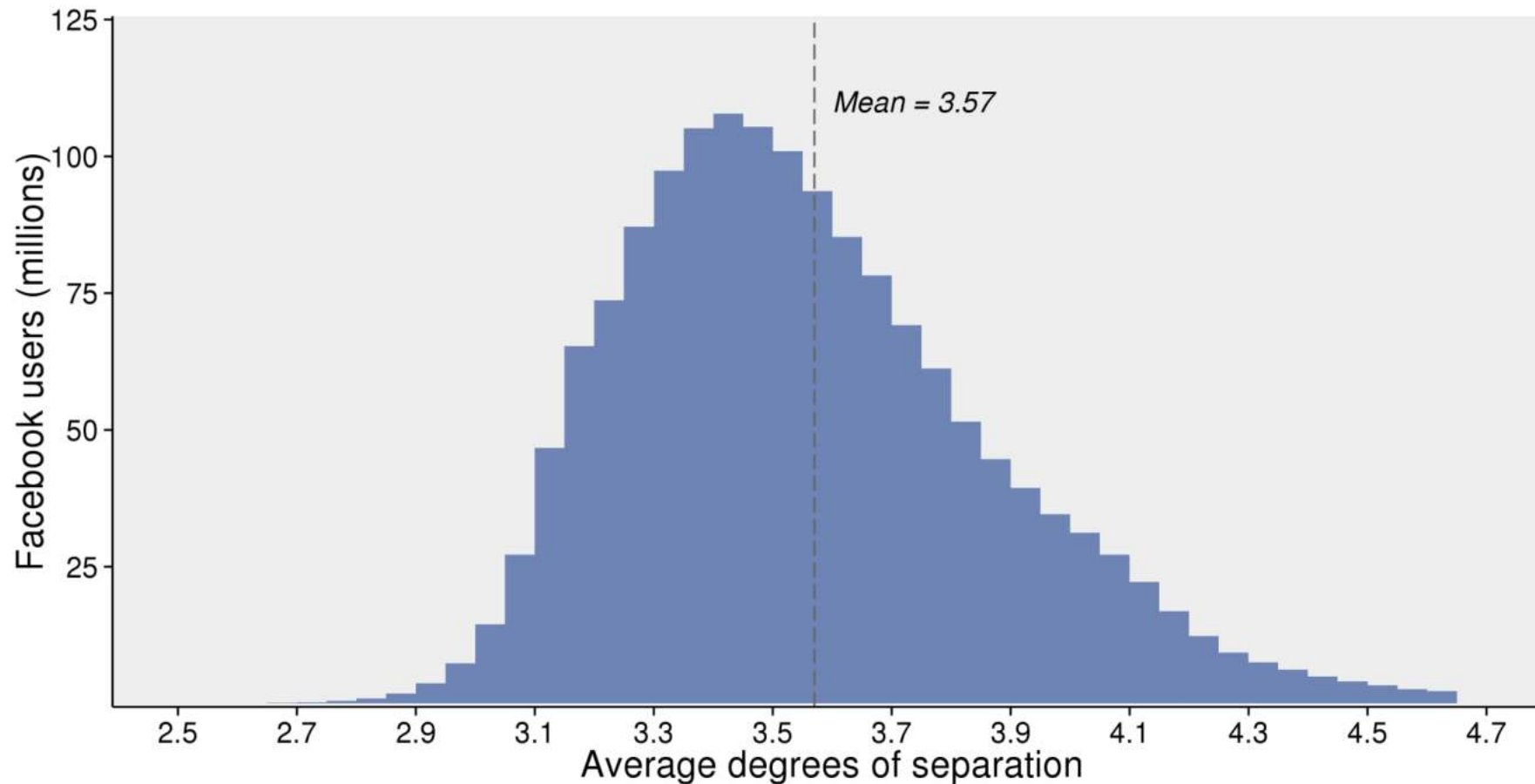
CREATING GROWTH, ENHANCING LIVES

Even for moderately sized networks, number of relationships scale (very) quickly

Degrees of separation on Facebook

In how many hops can you reach any other person on the FB platform

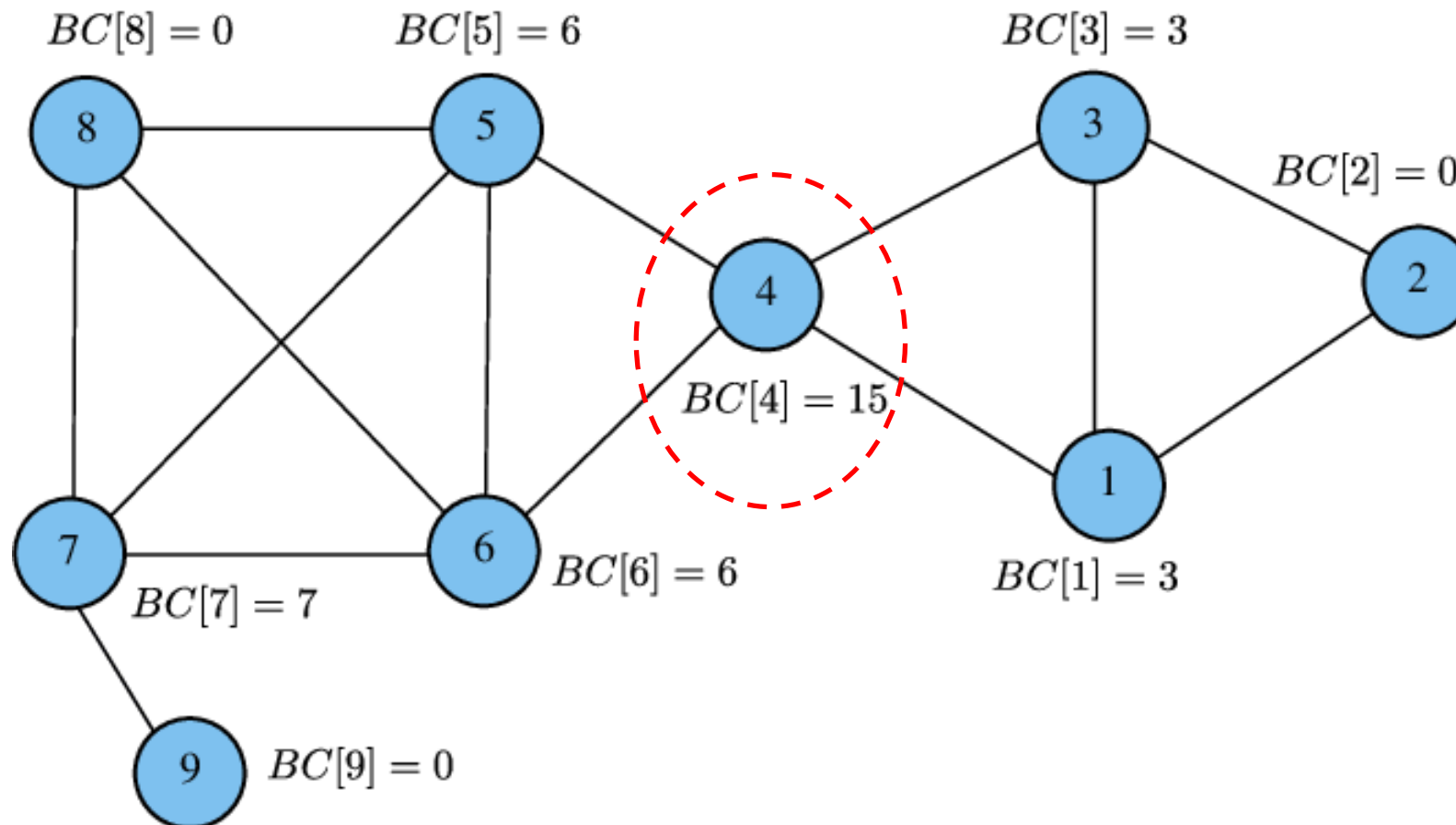https://research.fb.com/three-and-a-half-degrees-of-separation/

Computing graph distances at scale is both a HPC and a statistical challenge

- Imagine the network has 100 users

- Each user has 100 friends

- Each of these friends have 100 friends (i.e. each user has 100*100 friends-of-friends, and so forth)
  - Some overlap in friends, so actual number will likely be less than this

- Repeating shortest-path calculation for every pair of >2 billion users imposes huge computational burden

Computing certain network features also requires HPC

Betweenness centrality of X measures the proportion of all shortest paths that include the focal node X

$$BC[8] = 0 \quad BC[5] = 6 \quad BC[3] = 3$$
$$BC[2] = 0$$
$$BC[4] = 15$$
$$BC[7] = 7 \quad BC[6] = 6 \quad BC[1] = 3$$
$$BC[9] = 0$$

McLaughlin, A., & Bader, D. A. (2014, November). Scalable and high performance betweenness centrality on the GPU. In *SC'14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (pp. 572-583). IEEE.

High-value HPC use cases in finance & education that leverage network science

CREATING GROWTH, ENHANCING LIVES

Network science & creditworthiness analysis

Two motivating problems

1. Predicting **loan repayment and defaults** using network features particularly in emerging markets

   - Many emerging markets lack established credit infrastructure
   - Such markets have credit practices that are human intensive and bias-prone
   - Emergence of big data-driven credit scoring is transforming the market, but has challenges
   - In absence of financial history, social networks hold key signals but **computing "strong" network predictors is a HPC problem**

2. Understanding **contagion in good/bad repayment practices** within the network?

   - Is there evidence of peer-influence in loan repayment?
   - Can we infer/estimate cascades in loan default before they happen? **These involve simulation and estimation processes that are very computationally expensive!**
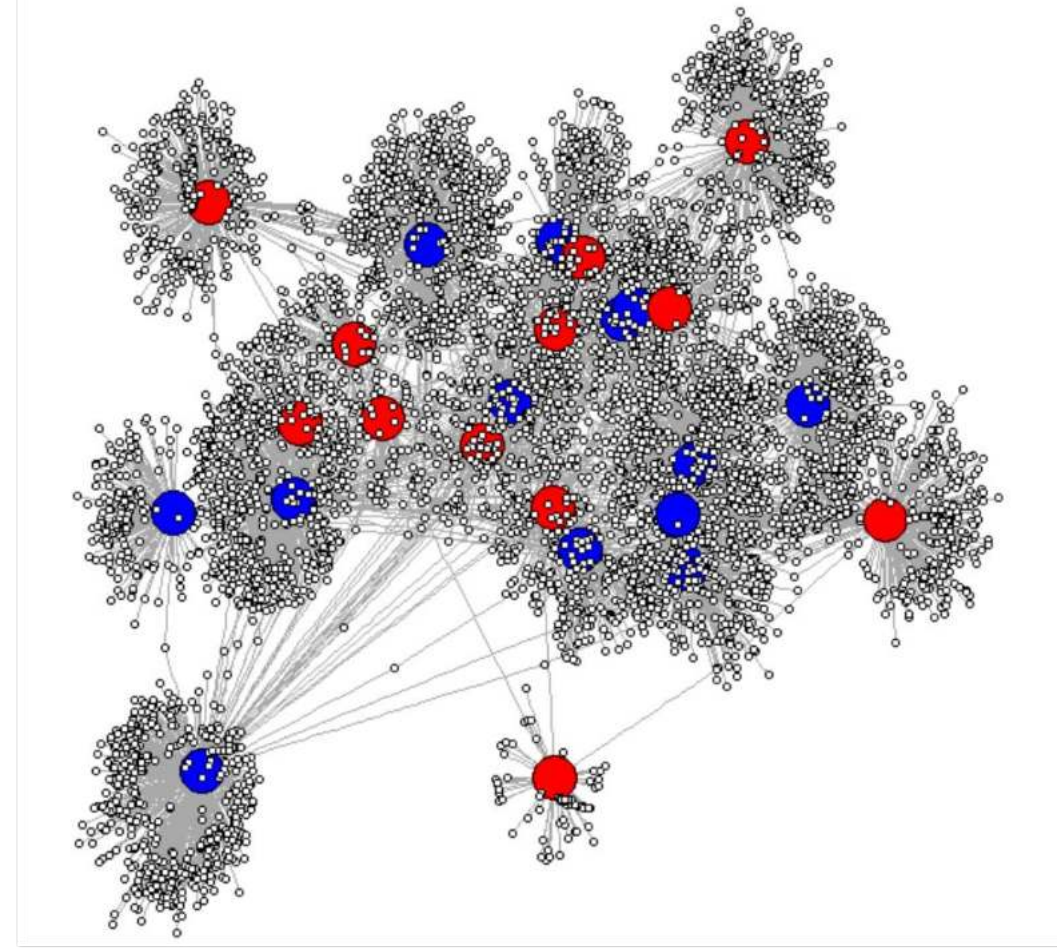   - Can we intervene early to prevent these cascades?

CREATING GROWTH, ENHANCING LIVES
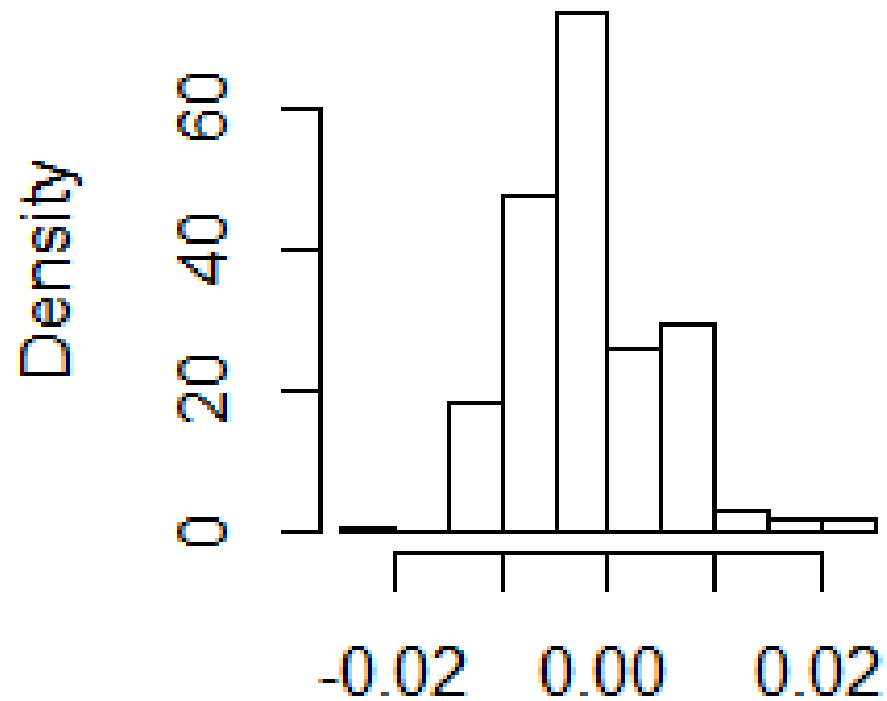
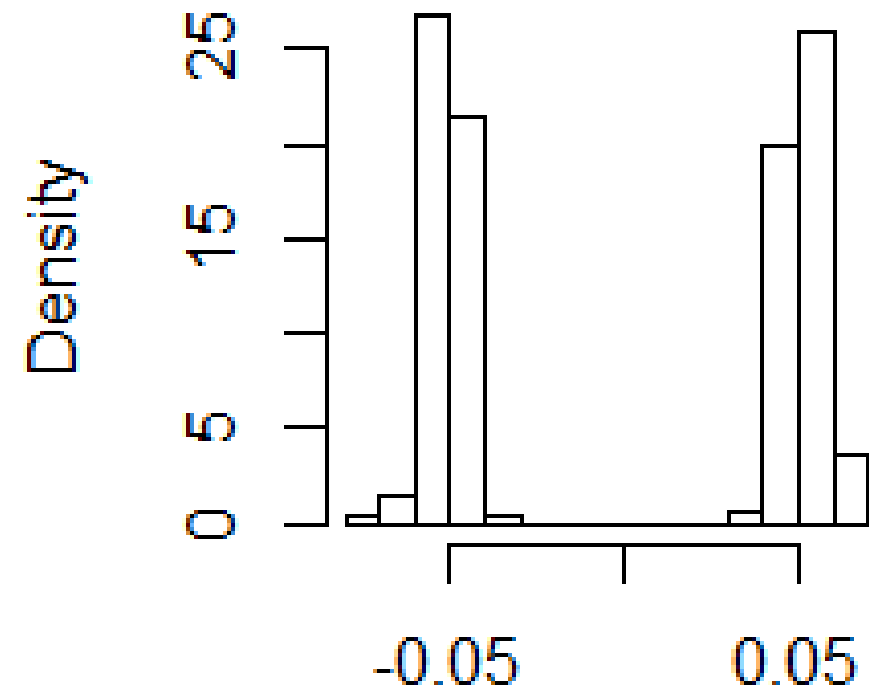# Illustrative example from an Asian city
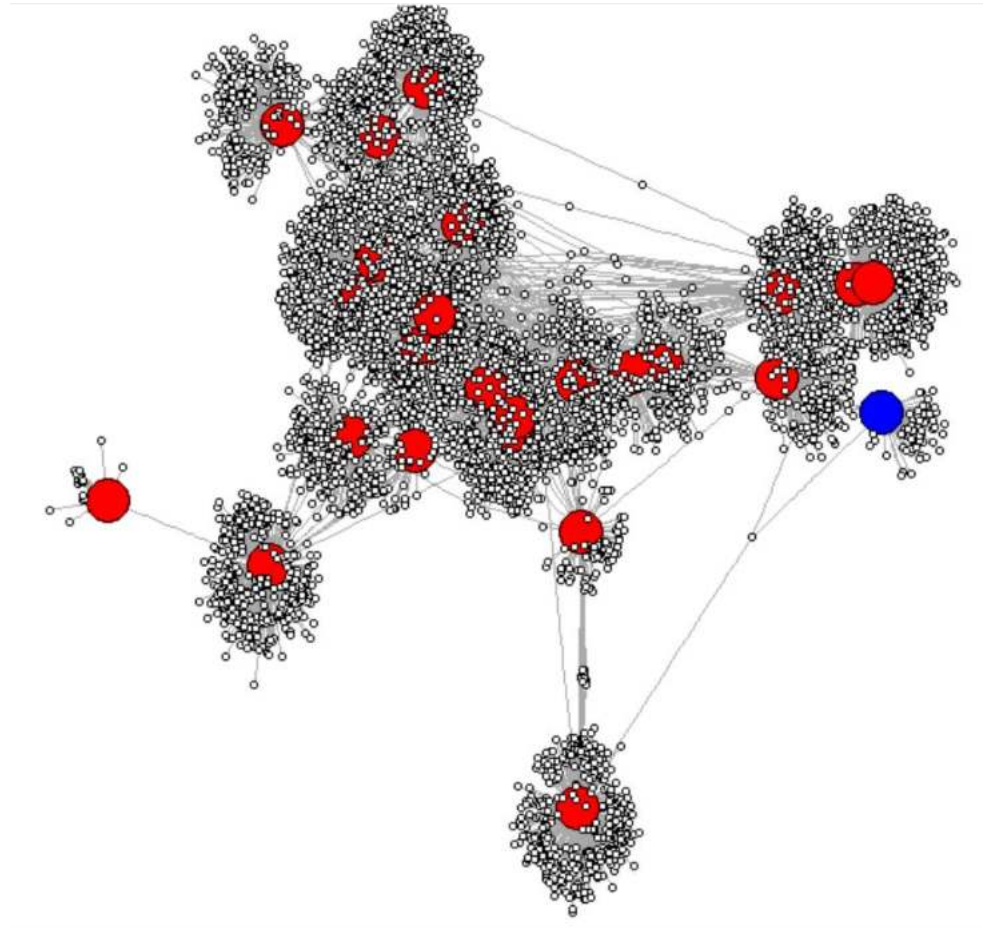
Low credit risk sample

High credit risk sample

Peer-influence varies across credit risk groups

Peer-influence in LOW credit risk sample

Peer-influence in HIGH credit risk sample

Network science & academic outcomes

Two motivating problems

1. Estimating impact of **online student networks** on academic performance & job outcomes

   - Social networks have historically been leveraged to achieve academic and employment outcomes

   - For the first time, we have an opportunity to quantify some of these effects using large-scale and fine-grained social interaction data

   - Solving this causal inference problem on networks requires experimental as well as **computational/HPC innovations**

2. Understanding **impact of COVID-19** in disrupting/reshaping student networks

   - There has been an increasing push towards online interactions, with limited physical interactions

   - Analyzing how these online interactions lead to network formation, and how they sustain over time is a key sociological and policy question

   - Can these online networks, in turn, drive offline interactions in future?

CREATING GROWTH, ENHANCING LIVES

17

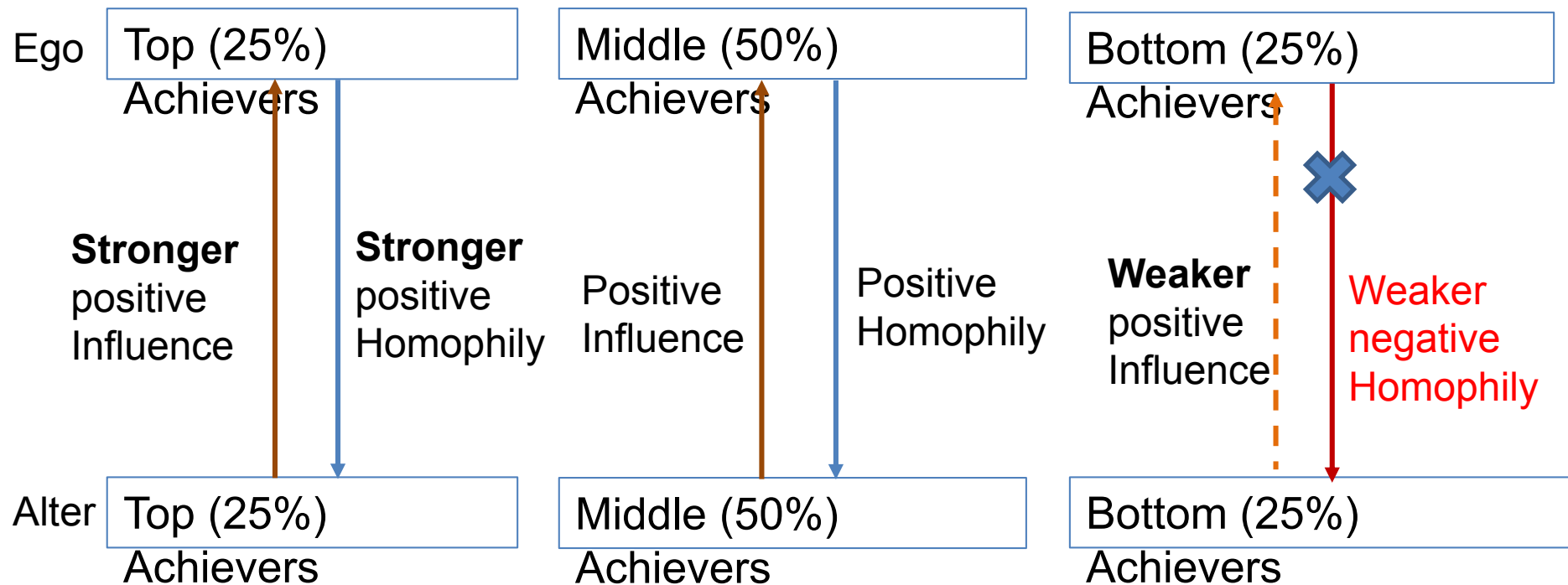# Online network interactions & academic outcomes
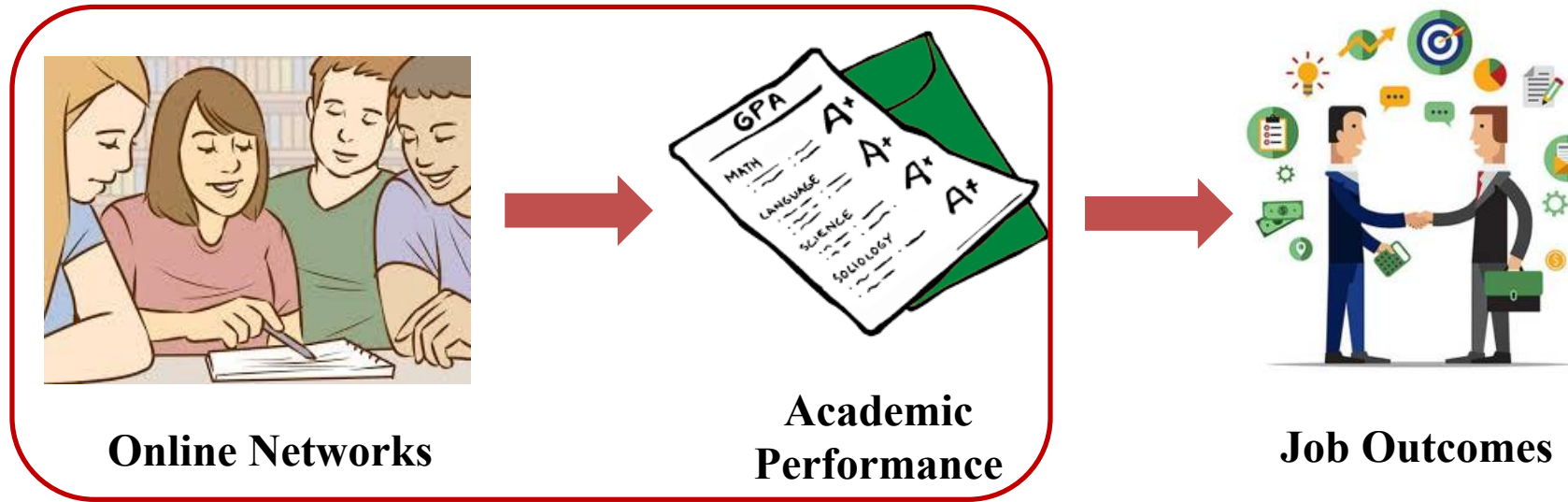
1. Do students form ties based on homophily in academic performance?

2. Does peer influence have a significant impact on academic performance?

3. Do these peer effects vary with students' academic abilities?

Ding, D., Bhattacharya, P., Phan, T. Q., & Lu, X. (2018). The Impact of Peer Influence on Academic Performance: A Three-Stage Co-Evolution Model. ICIS 2018

Student network is increasingly clustered by grades

T=1   T=2   T=3   T=4

T=5   T=6   T=7   T=8

GPA=1   GPA=2   GPA=3   GPA=4   GPA=5
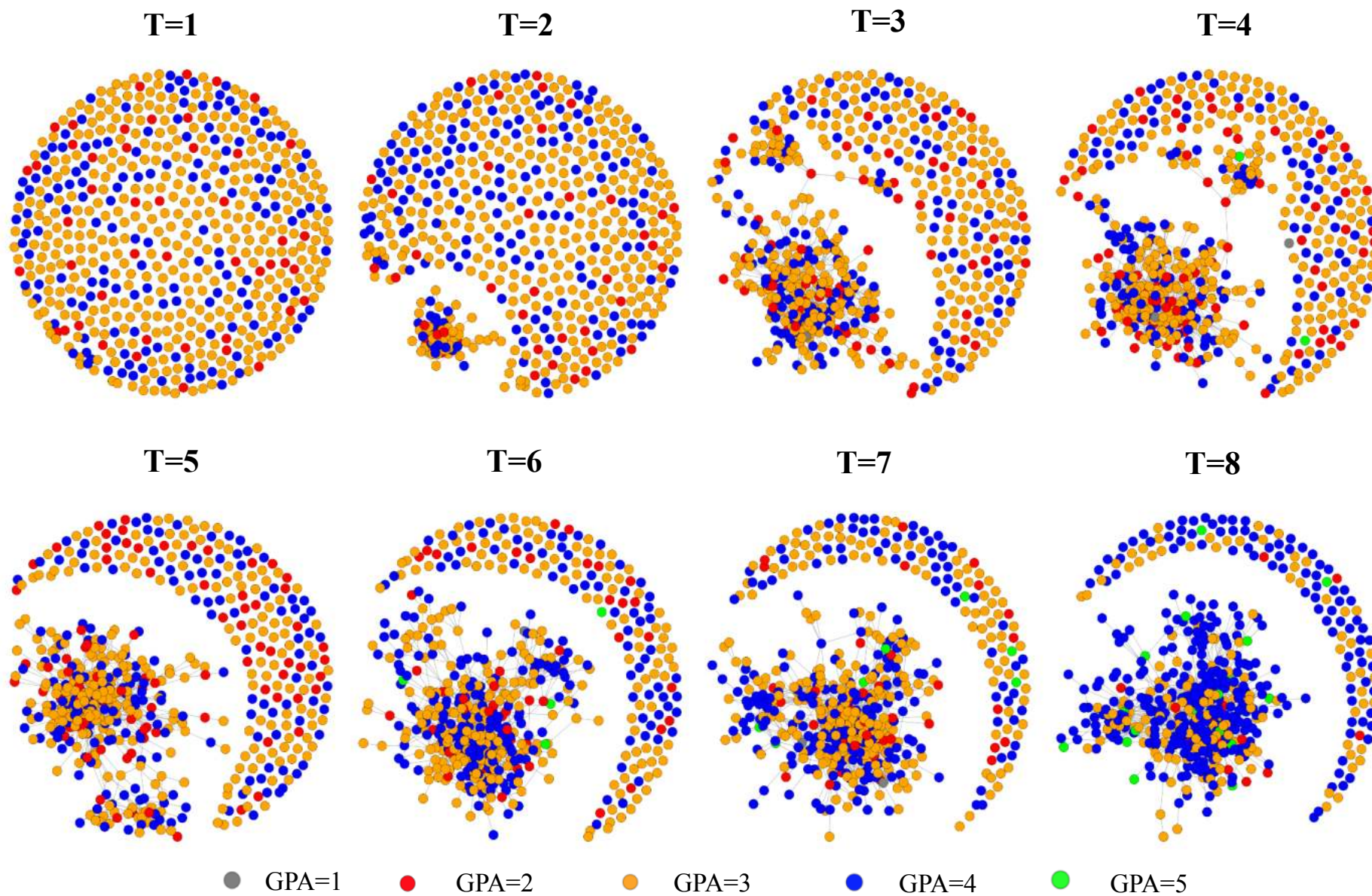
Estimation results from a Stochastic Actor-oriented Model (SAOM)

Heterogeneity in homophily as well as peer-influence

Key Takeaways

1. Network science has the potential to significantly transform existing businesses and operating models

2. The increasing availability of big data in domains like finance, healthcare and education has opened up avenues for studying large and real-word networks

3. Analyzing such large networks is at once a HPC and a statistical challenge because of the scale in size and relationships
   - With the right combination of HPC infrastructure and talent this can be a tremendous opportunity

CREATING GROWTH, ENHANCING LIVES

THANK YOU !

Dr. Prasanta BHATTACHARYA

prasantab@ihpc.a-star.edu.sg

Sept 23, 2020

Image source: https://www.flickr.com/photos/33479292@N05/3536127129//