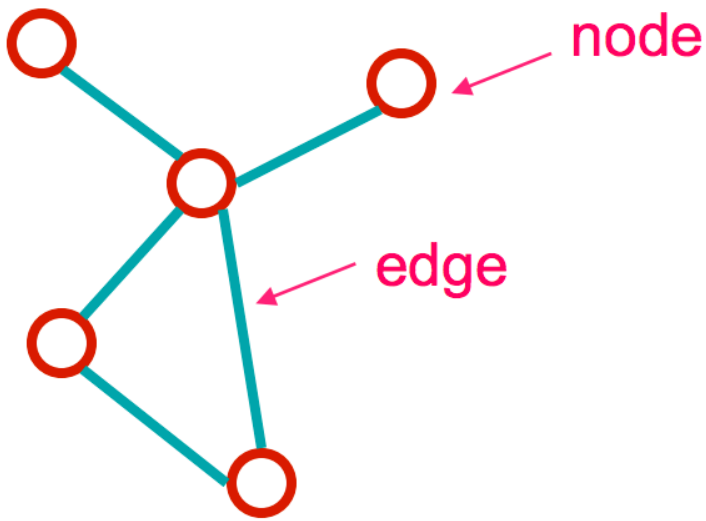
Facebook
Tree-like Structure in Social and Information Networks

Dustin Tran, Cole Diamond
Institute for Applied Computational Science @
Harvard University

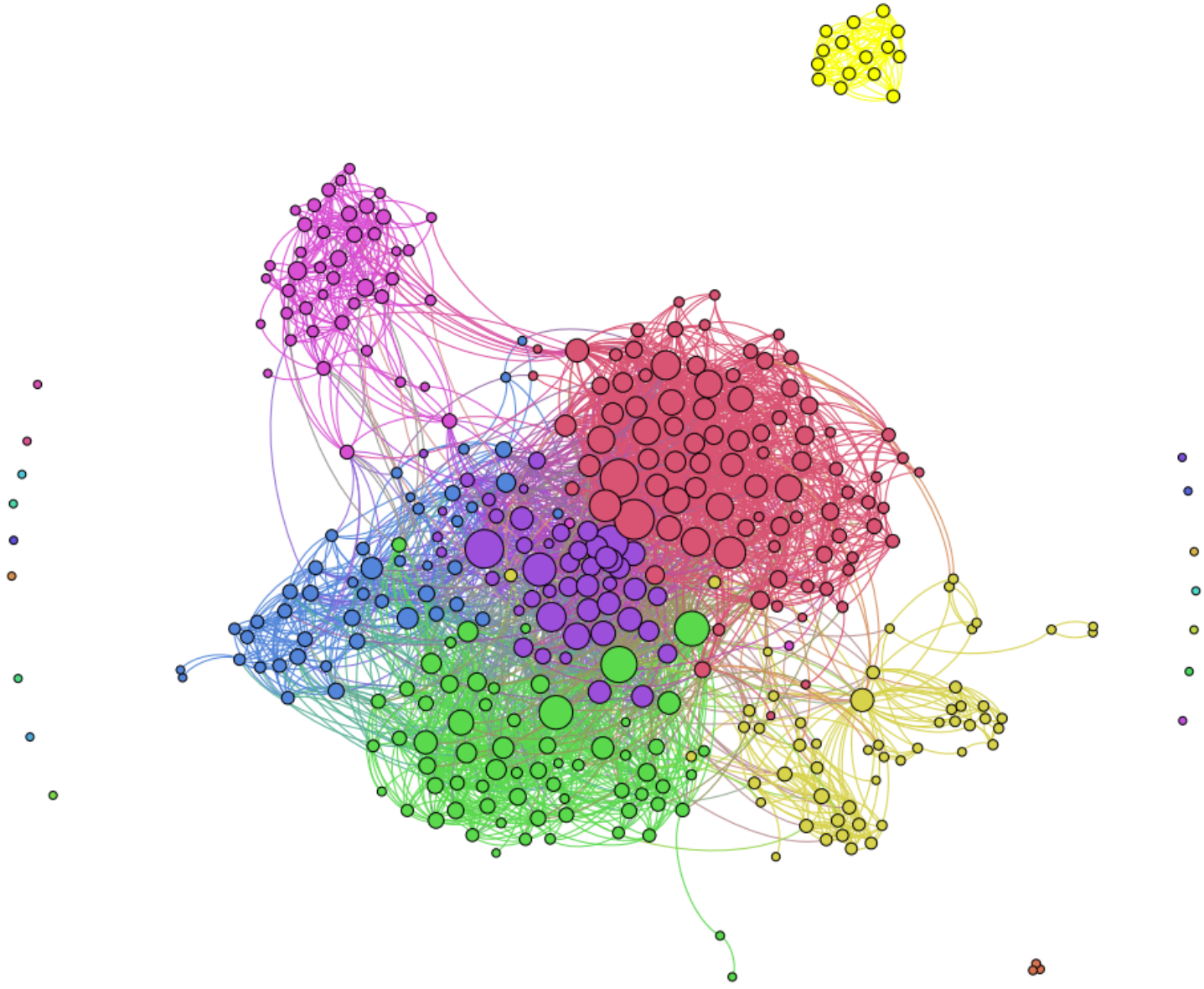
Fundamentals

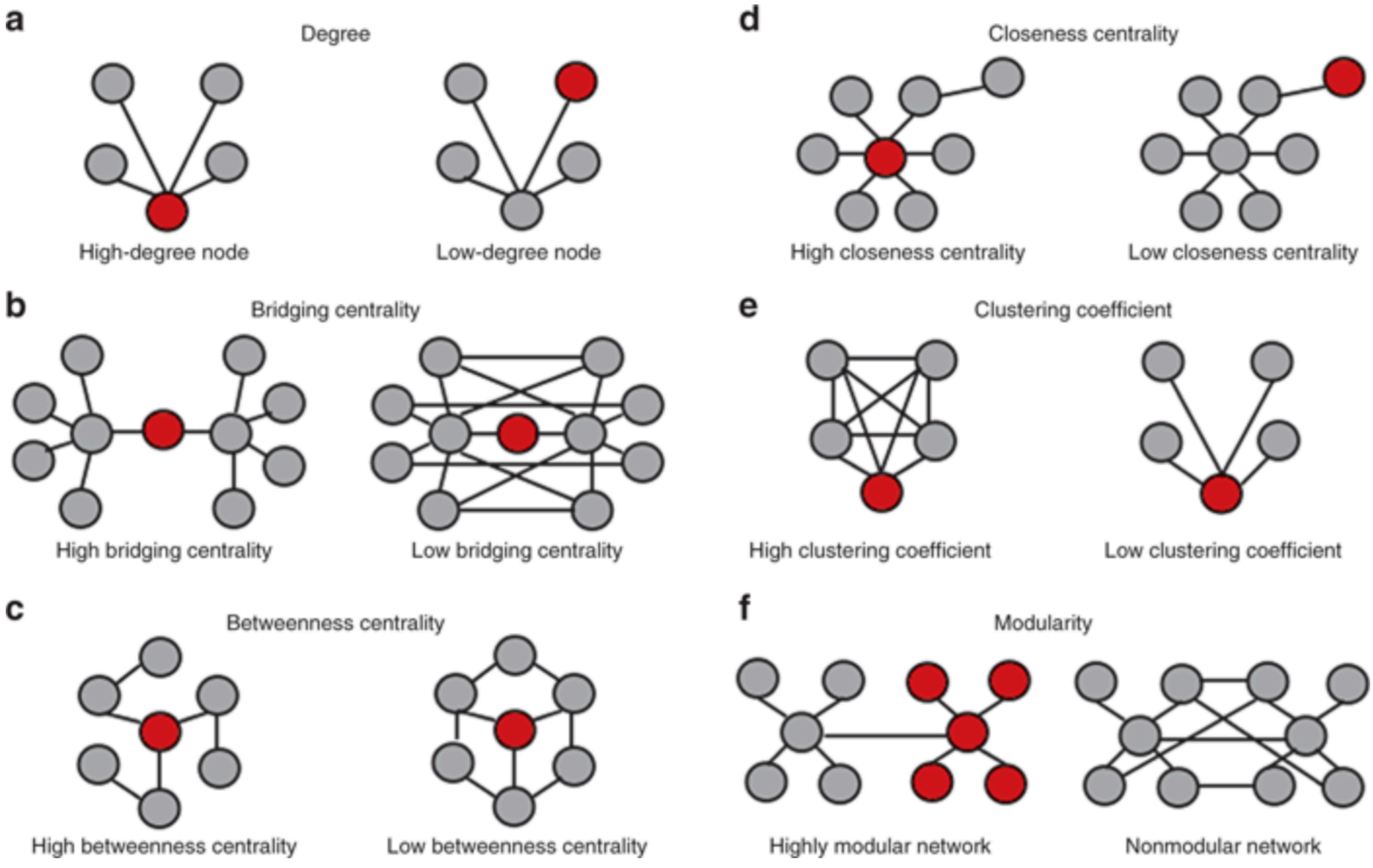
A *network* is a pair (V, E) where V is a set whose elements are known as *vertices* and E is the set of ordered pairs of vertices known as *edges*.



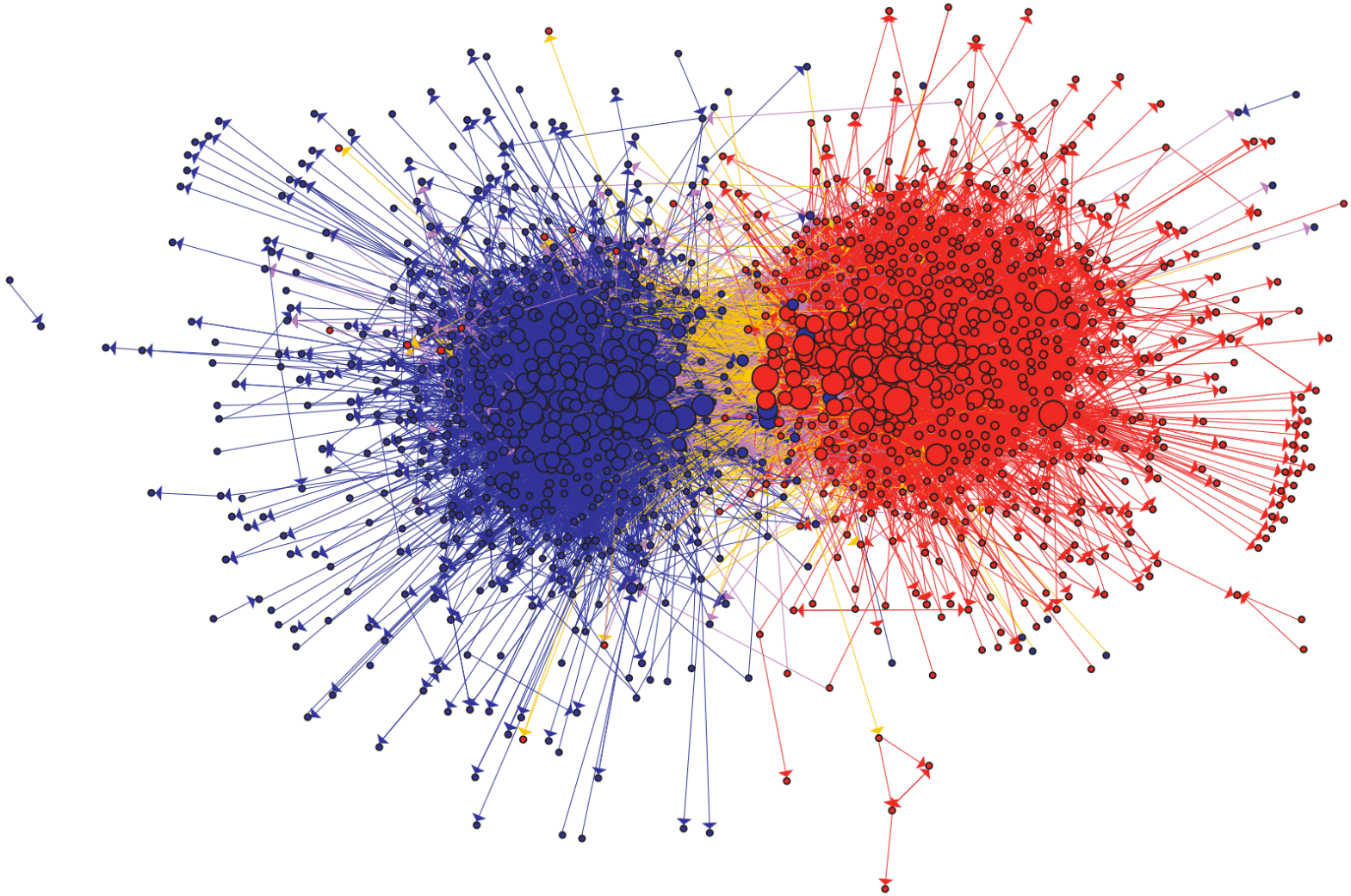
points	lines	
vertices	edges, arcs	math
nodes	links	computer science
sites	bonds	physics
actors	ties, relations	sociology

Social Network

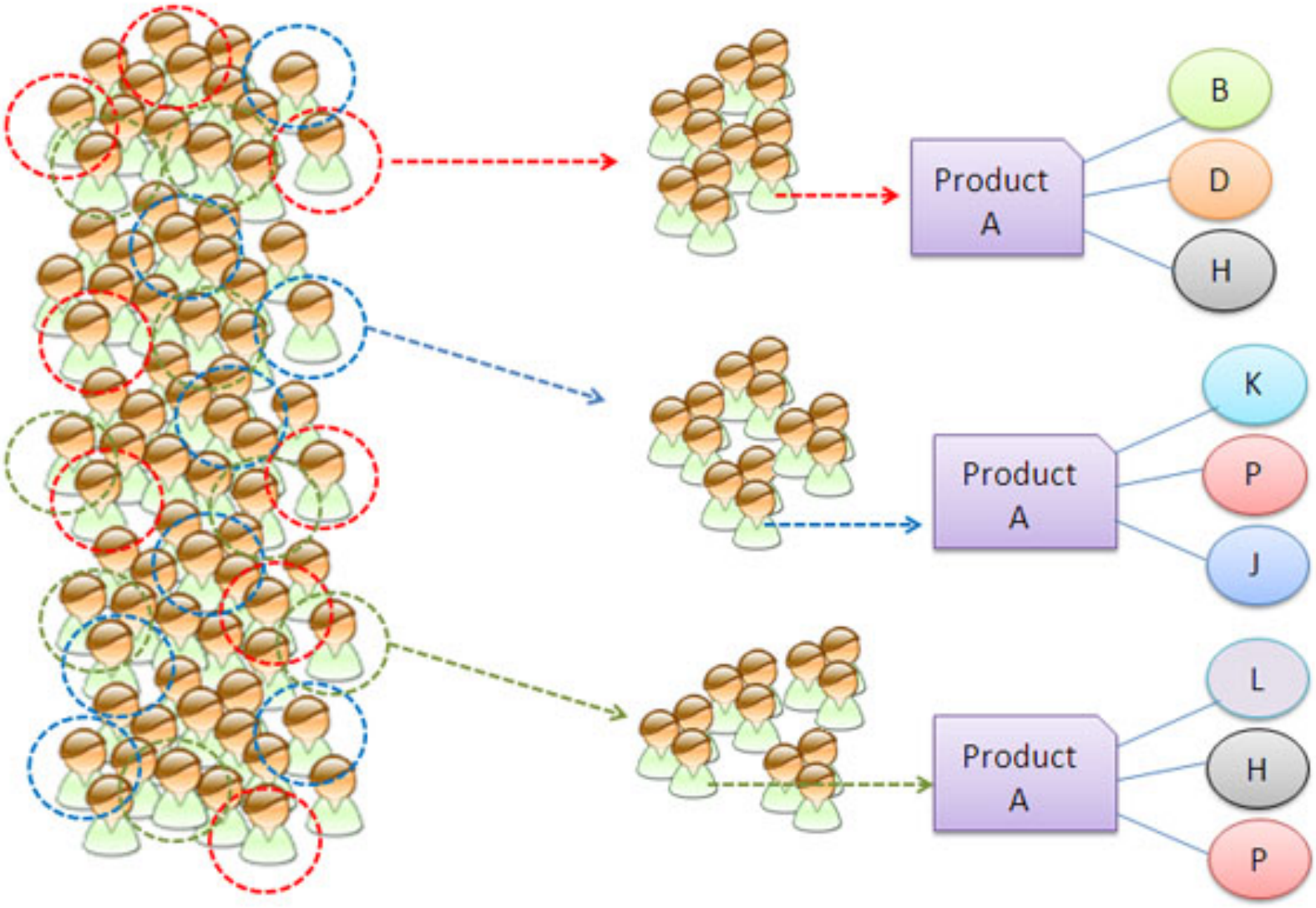



a
Degree
High-degree node
Low-degree node
b
Bridging centrality
High bridging centrality
Low bridging centrality
c
Betweenness centrality
High betweenness centrality
Low betweenness centrality
d
Closeness centrality
High closeness centrality
Low closeness centrality
e
Clustering coefficient
High clustering coefficient
Low clustering coefficient
f
Modularity
Highly modular network
Nonmodular network

Applications: Community Detection



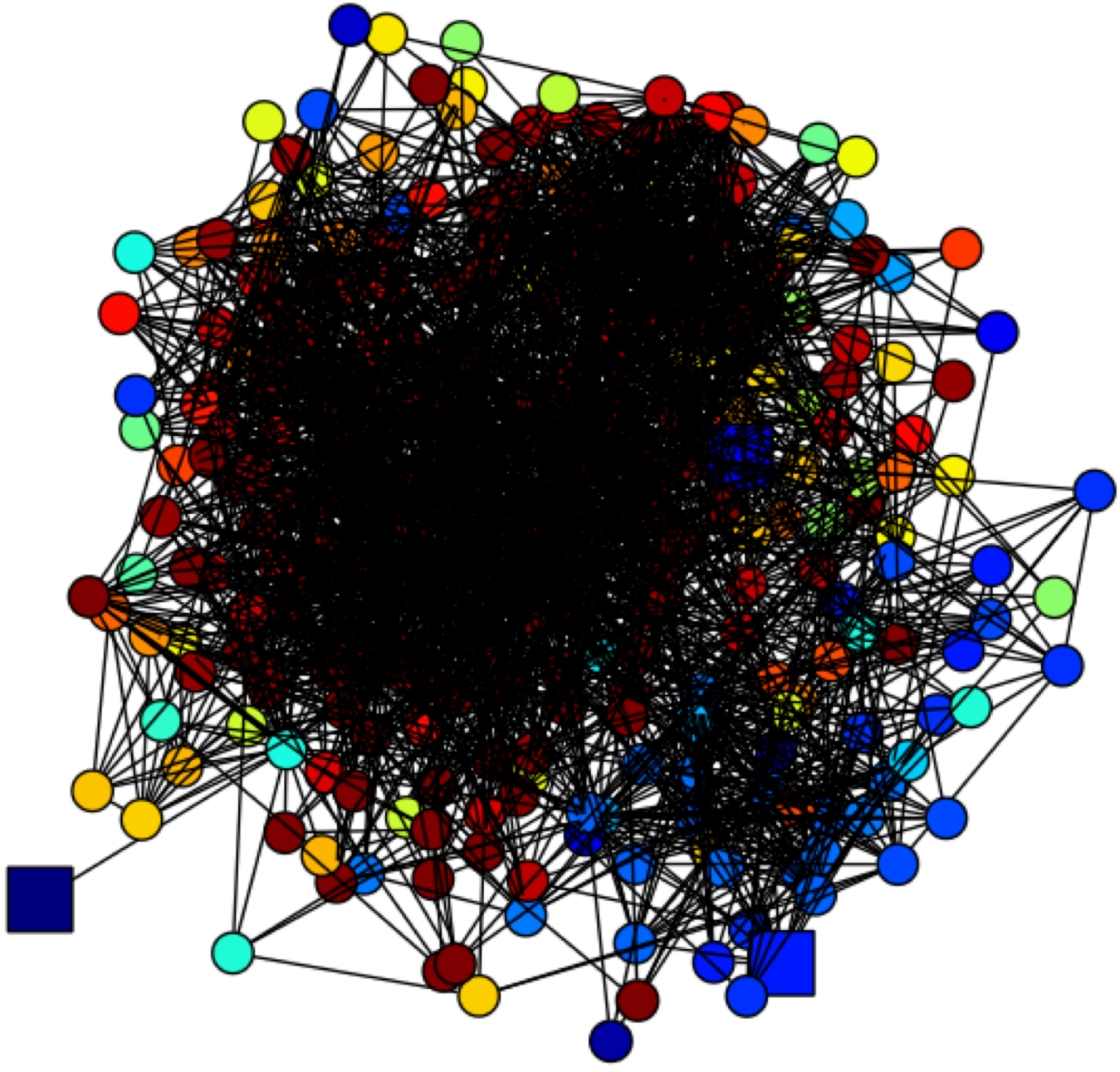
Applications: Collaborative Filtering



Applications: Information Propagation

- Viruses in computer networks
- Disease and vaccine spread
- Memes, viral marketing, gossip
- Innovation

Applications: Information Propagation



Complex Networks

Key characteristics:

- Small diameters
- Degree heterogeneity
- Exhibit clustering

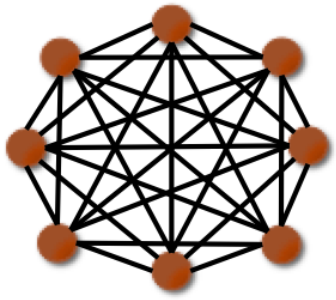
Simple models only capture some of these.

- *Erdös-Rényi*: Small diameter (+), Poisson degrees (-), No clustering (-)
- *Power law configuration graph*: Small diameter (+), Degree heterogeneity (-), No clustering (-)

Small Diameters

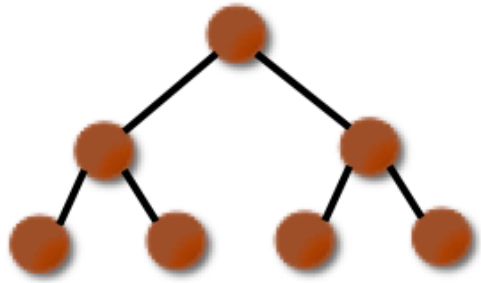
Clique

Constant



d-tree

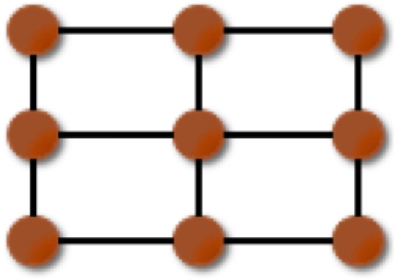
$\log(n)$



Large Diameters

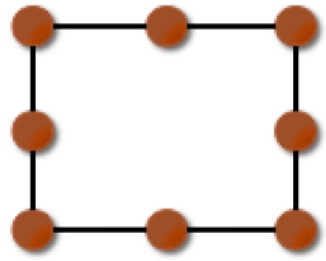
$n \times n$ planar grid

$\sqrt{n}$

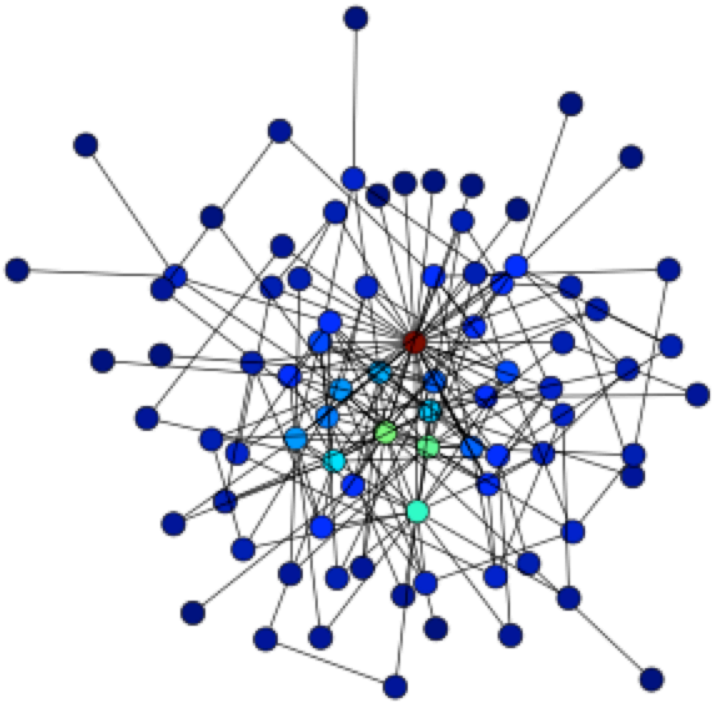


n-cycle

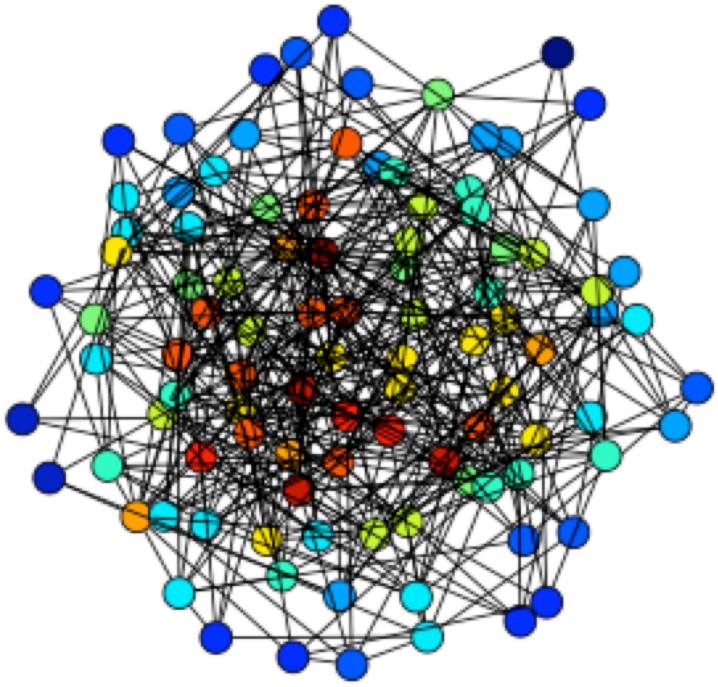
$\frac{n}{2}$



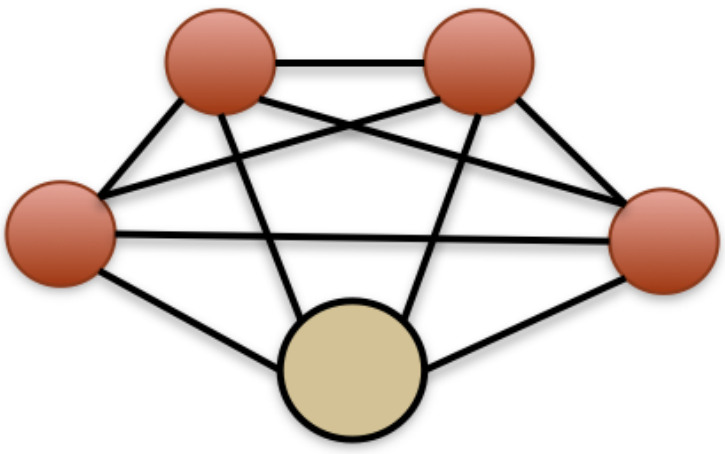
More heterogeneity
Max: 41, Min: 1



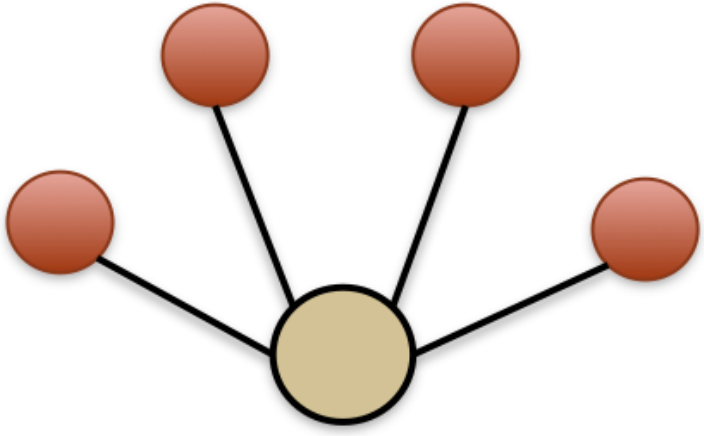
Less heterogeneity
Max: 17, Min: 3



Maximal
Clustering: 1.0



Minimal
Clustering: 0.0

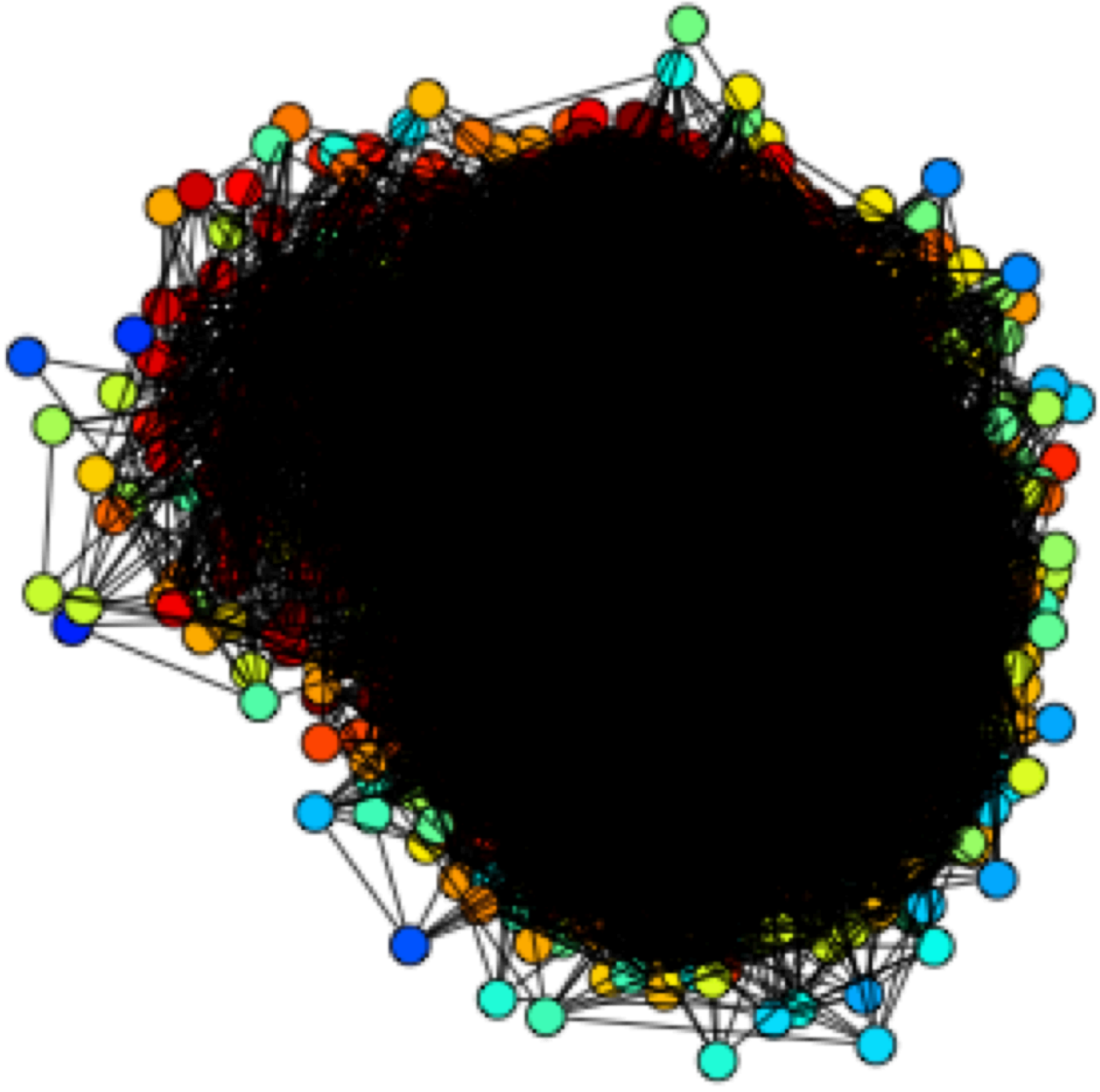


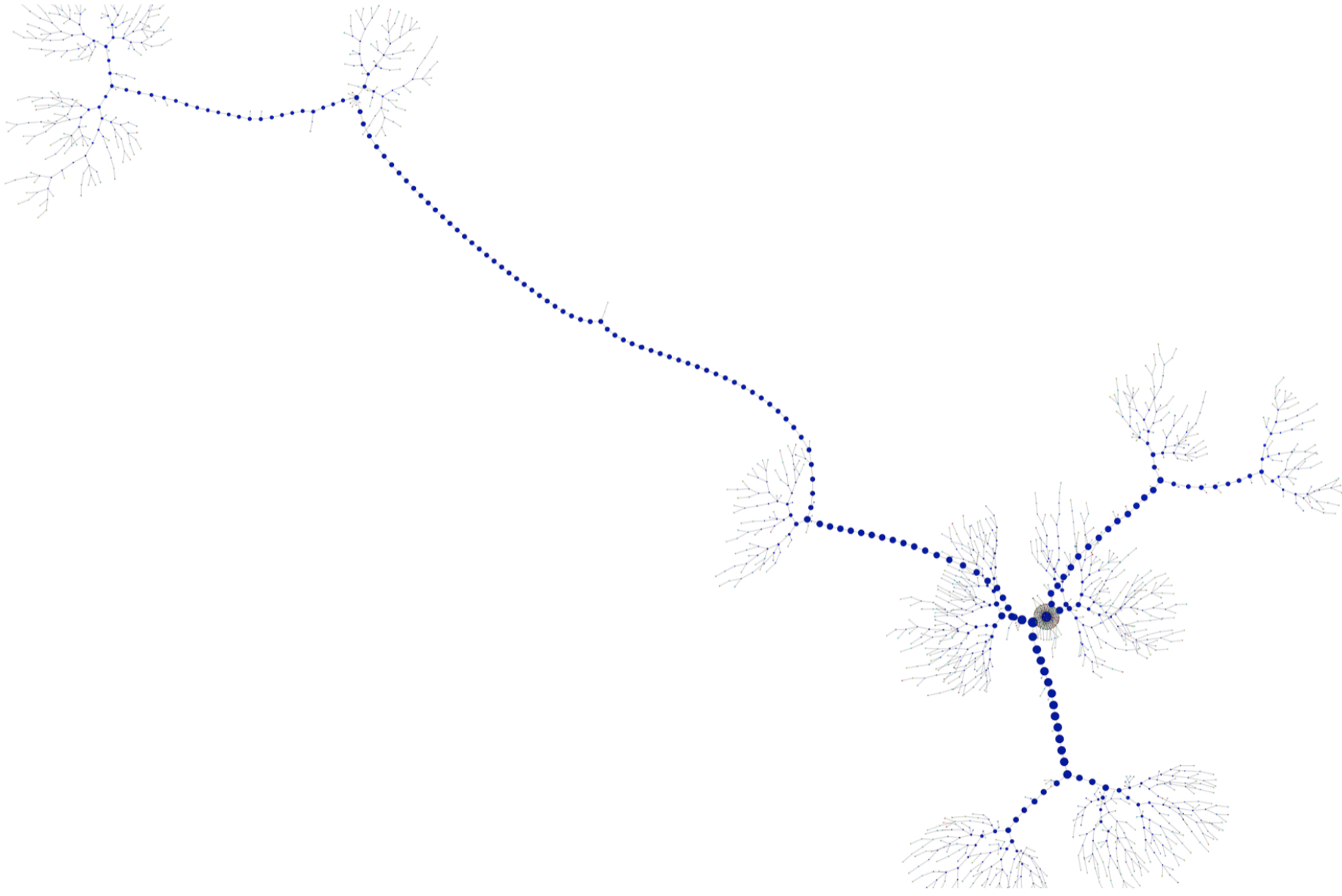
Why tree decompositions?

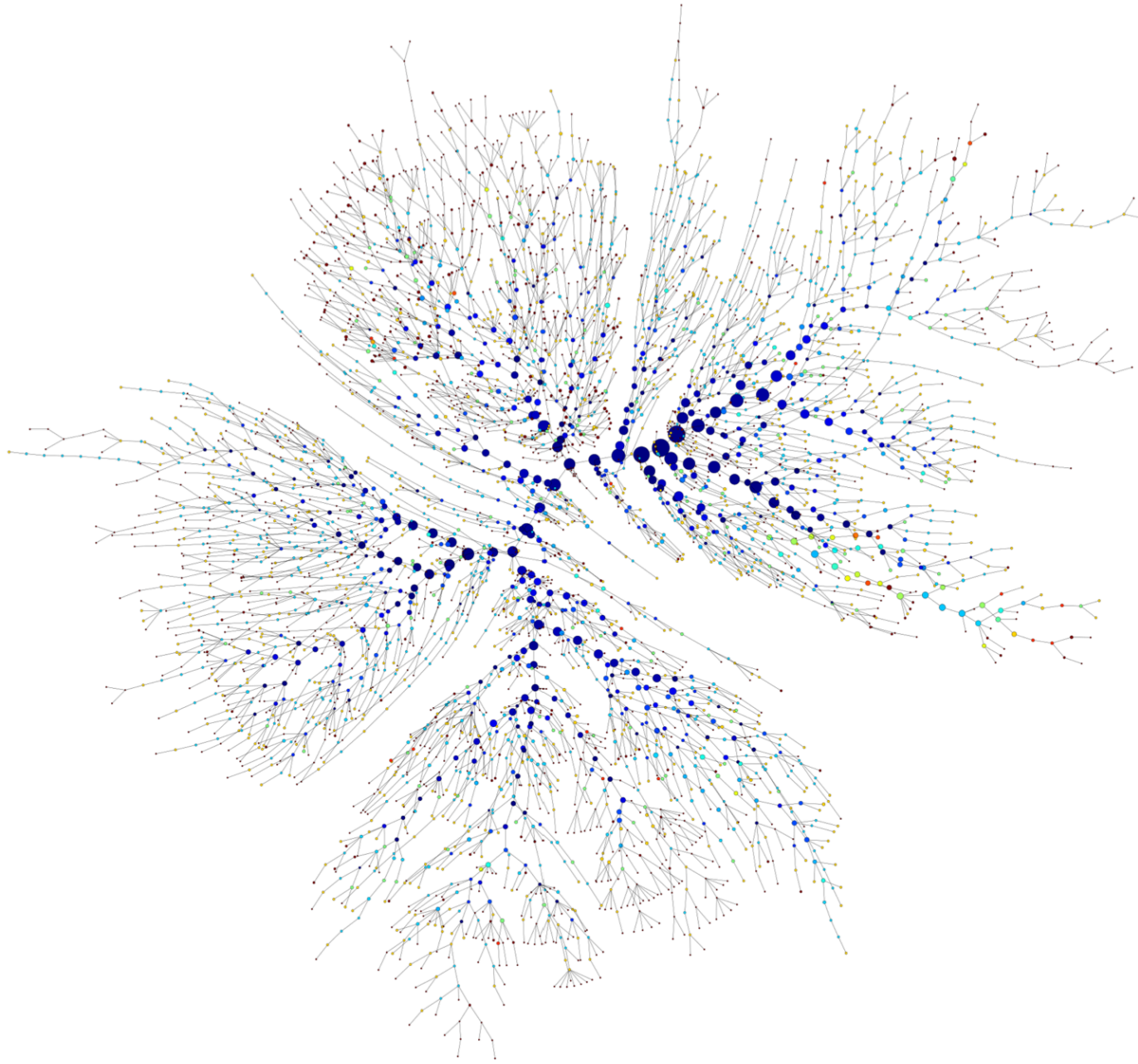
- Regular trees have *small diameter*, and assuming the network is hyperbolic and geometric, it preserves *degree heterogeneity* and is *well-clustered*

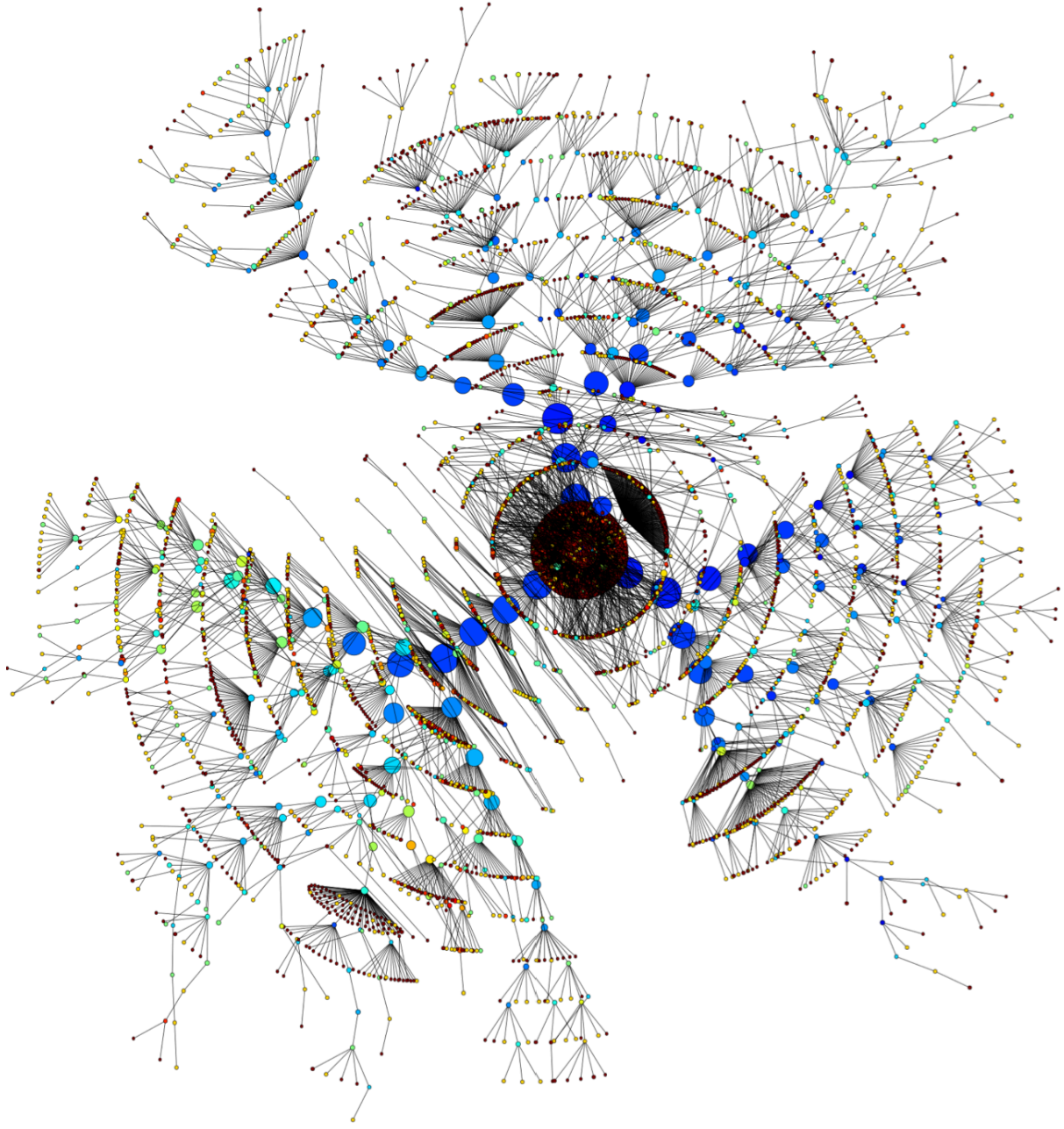
Can apply network models more easily

- Tree-based *belief propagation* (junction tree algorithm) over ”loopy” belief propagation
- Lesser need for *variational methods* or MCMC
- Computationally easier *graphical models* in general (deep belief networks, Markov random fields, etc.)









YASIV

Your Facebook Universe. Visualized.

See how your world is connected and discover new friends.

Visualize Friends

But please make sure your browser supports WebGL. Check here: get.webgl.org

Eric Berlow and Sean Gourley:

Mapping ideas worth spreading

TED2013 · 7:55 · **Filmed** Feb 2013
Subtitles available in 31 languages

View interactive transcript

References

- Adcock, A.; Sullivan, B.; Mahoney M. *Tree-like structure in large social and information networks.* ICDM, pages 1–10, 2013.
- Goldenberg, A.; Zheng, A.; Fienberg, S.; Airoldi, E. *A survey of statistical network models.* Foundations and Trends in Machine Learning, 2(2):1-117, 2009.
- Leskovec, J.; Lang, K.; Dasgupta, A.; Mahoney, M. *Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters.* CoRR, 2008.

@dustinvtran • dustinvtran.com