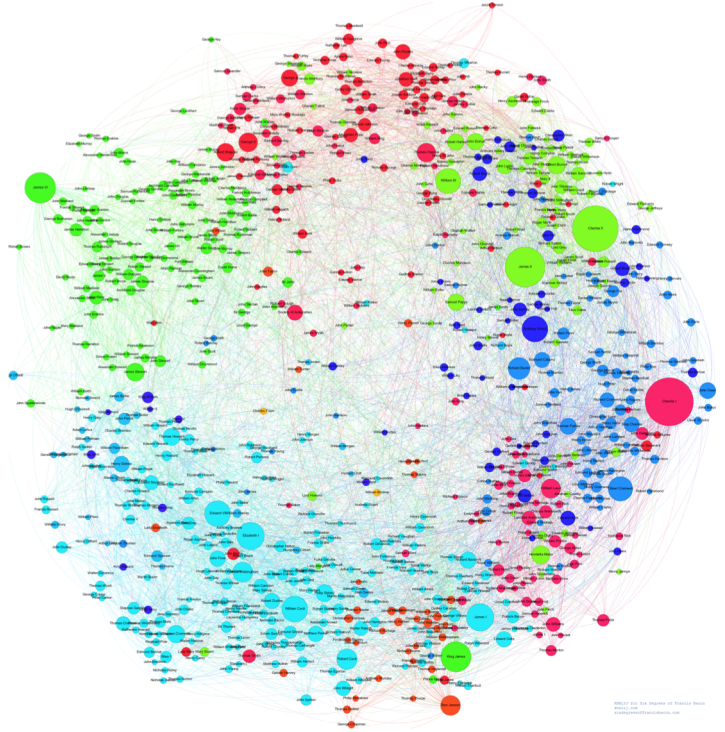


Influence Maximization on Social Networks

The Seed Selection Problem

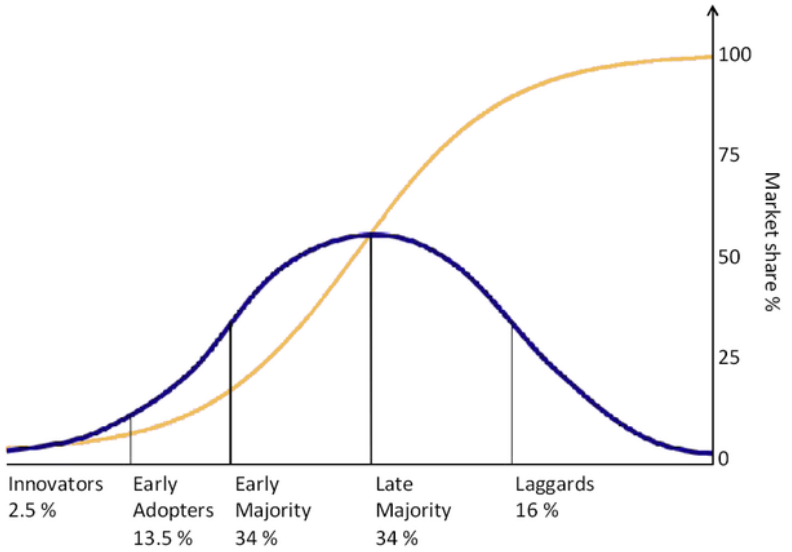
Gönenç Yücel

Boğaziçi University, Industrial Engineering Department

25 May 2018

Social diffusion processes

- Processes related to the spread of a **thing** within a social group
  - Spread of (fake) news/rumors
  - Influenza pandemics
  - Innovation diffusion
  - Diffusion of political views
  - Diffusion of behaviors



- In general terms, driven by three factors
  - Embodied properties of the diffusant
  - Individual state
    - condition, preferences, resources
  - Social influence
    - Information, influence

Influence diffusion and maximization

- Diffusion processes where the dominant factor is the social influence
  - Embodied properties are not important
  - Individual behavior is primarily conditioned by what others do

- Influence maximization
  - Given a social group, maximizing the final diffusion level (i.e. total influence) by selecting the initial set of individuals to adopt the diffusant (i.e. seed set)
    - A.k.a. The seed selection problem, target set selection problem

Modelling the diffusion process

- Social models of the adoption process at the individual level
- Linear Threshold (LT) Model
- Independent Cascade (IC) Model

Structures of social networks

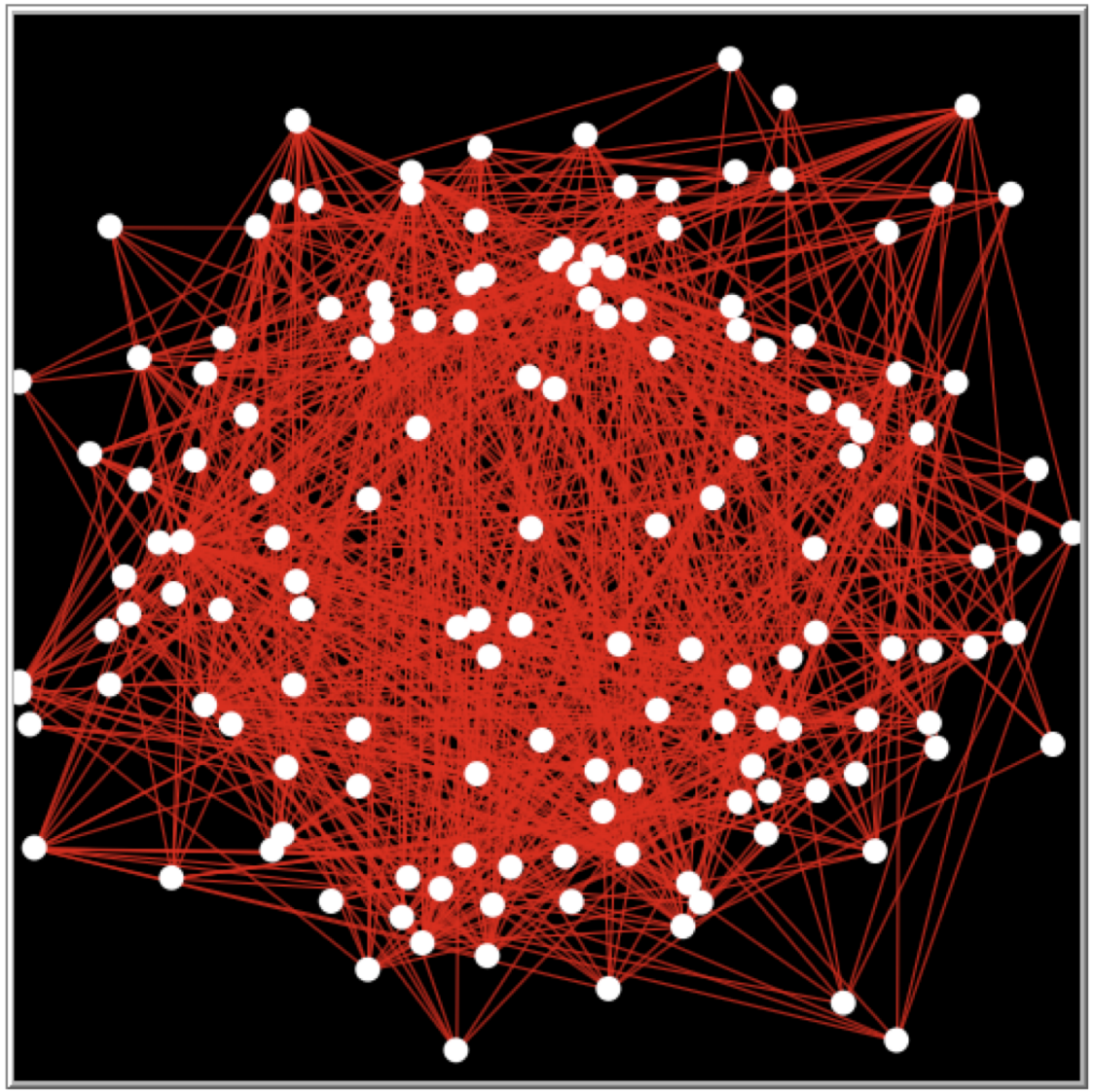
- Key structural features of a network;
  - # of nodes and edges
  - Diameter
  - Avg. shortest path length (ASPL)
  - Clustering coefficient (CC)
  - Degree distribution (DD)

- Three major classes of networks
  - Random networks
  - Scale-free networks (Power law in DD)
  - Small-world networks (High CC, Low ASPL)

		Solution Approach	
		Optimization by MP	Rule-based Heuristics
Network Information	Perfect and Complete	**A**	**B**
	Imperfect and/or Incomplete	**C**	**D**

Idea

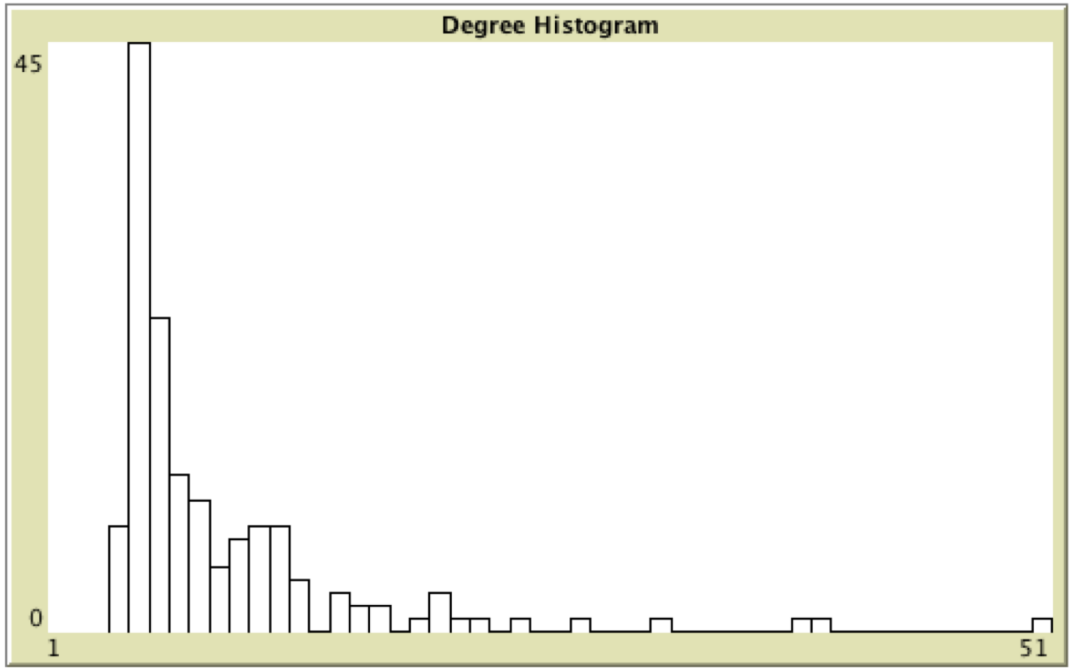
- Synthetically generate a large set of social networks of a certain class
  - Possess full and perfect information to simulate a diffusion
- Pretend that we do not know anything about the network and select seeds with candidate rules
- Evaluate the rules based on the resulting diffusion levels



Create Network | Layout | propagation_method: Independent Cascade | network_type: Scale-Free

Select Seeds | Seed Selection by Hand | heuristic: Sw1S

Reset Network | Propagate



Clustering Coefficient	Average Degree	Average Distance	Diameter
0.152	9.053	2.412	4

Experimental setup

- Linear Threshold Model
- Seed set budget: 5% of the network size
- 3 network classes
  - 20 heuristics to compare
  - 300 instances for each network
- Two groups of selection approaches (heuristics)
  - Group 1: Requires complete network information
  - Group 2: Rely on local information about randomly picked sections
- Initial analysis on 1000-node networks
  - Scale-up to 50000-node ones

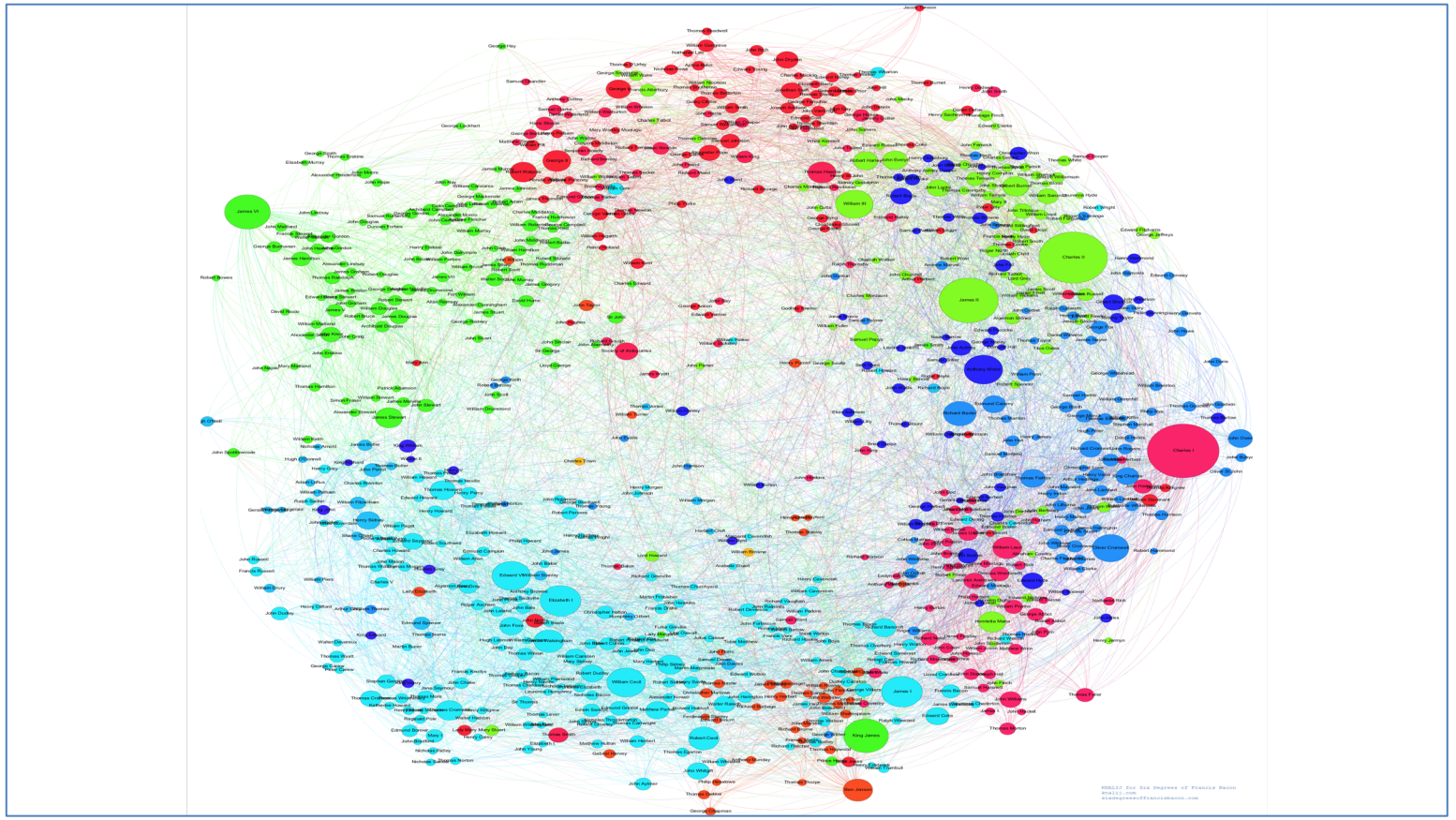
Sample selection-rule couple

- Average Threshold (AT):
  - Choose s nodes in the network with minimum average threshold metric to activate.
  - The average threshold for node x is calculated as the mean of the neighbor thresholds of node x.

- Average Threshold within 1 Step (ATw1S):
  - This heuristic first accesses to an inactive node.
  - Then, it selects the inactive neighbor of this node with the minimum average threshold value as a seed.
  - This procedure is repeated s times. In real life, this corresponds to asking someone how convincing are their friends.

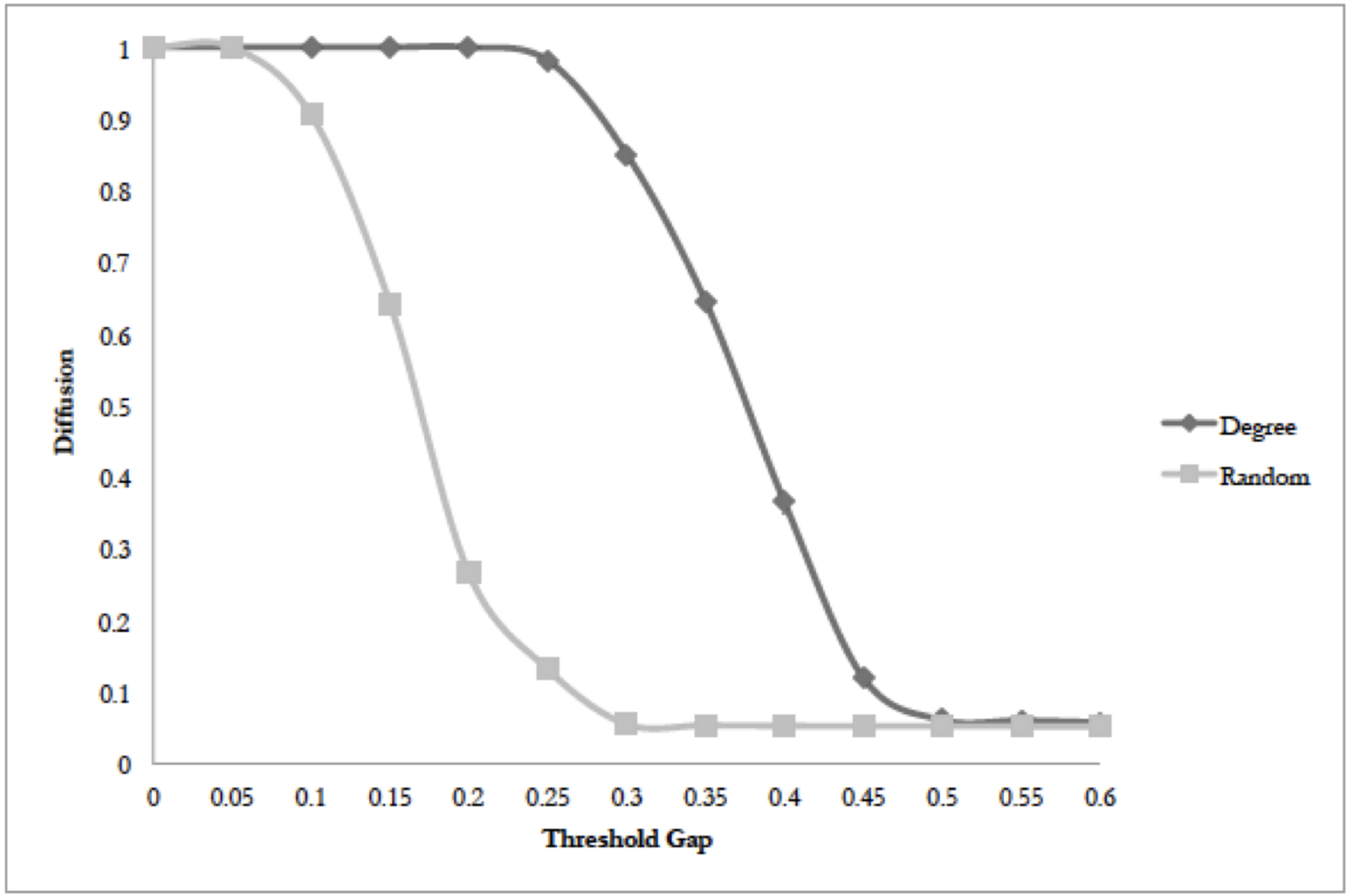
Conclusions

- No silver bullet.
  - Importance of apriori knowledge on the network class and the right diffusion model
- Value of full and perfect information
  - Full and perfect information may even be misleading in some cases
- Just knowing what kind of network we are dealing with vs knowing everything about that network
- Evolution of the network

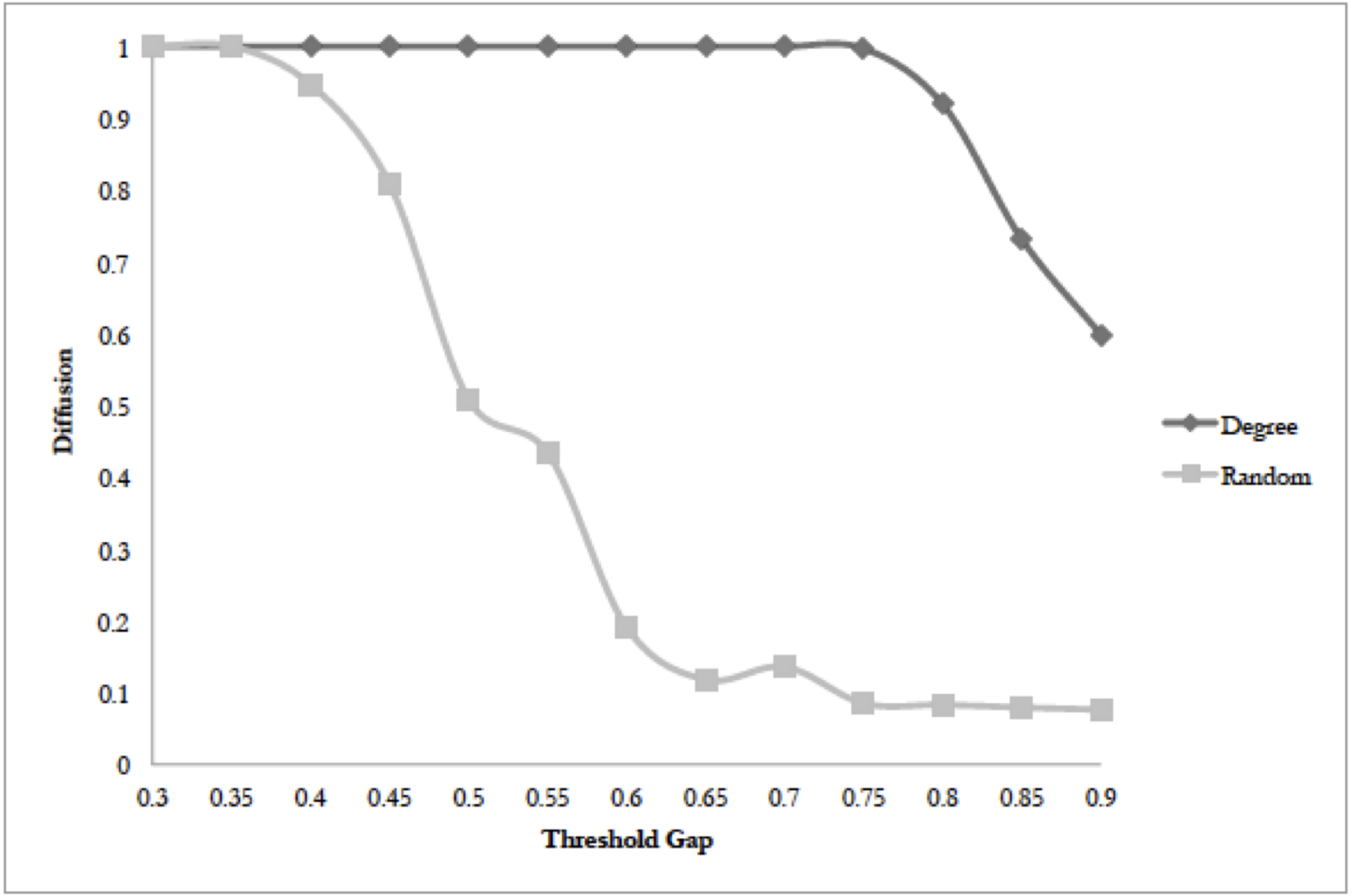


Thank you…

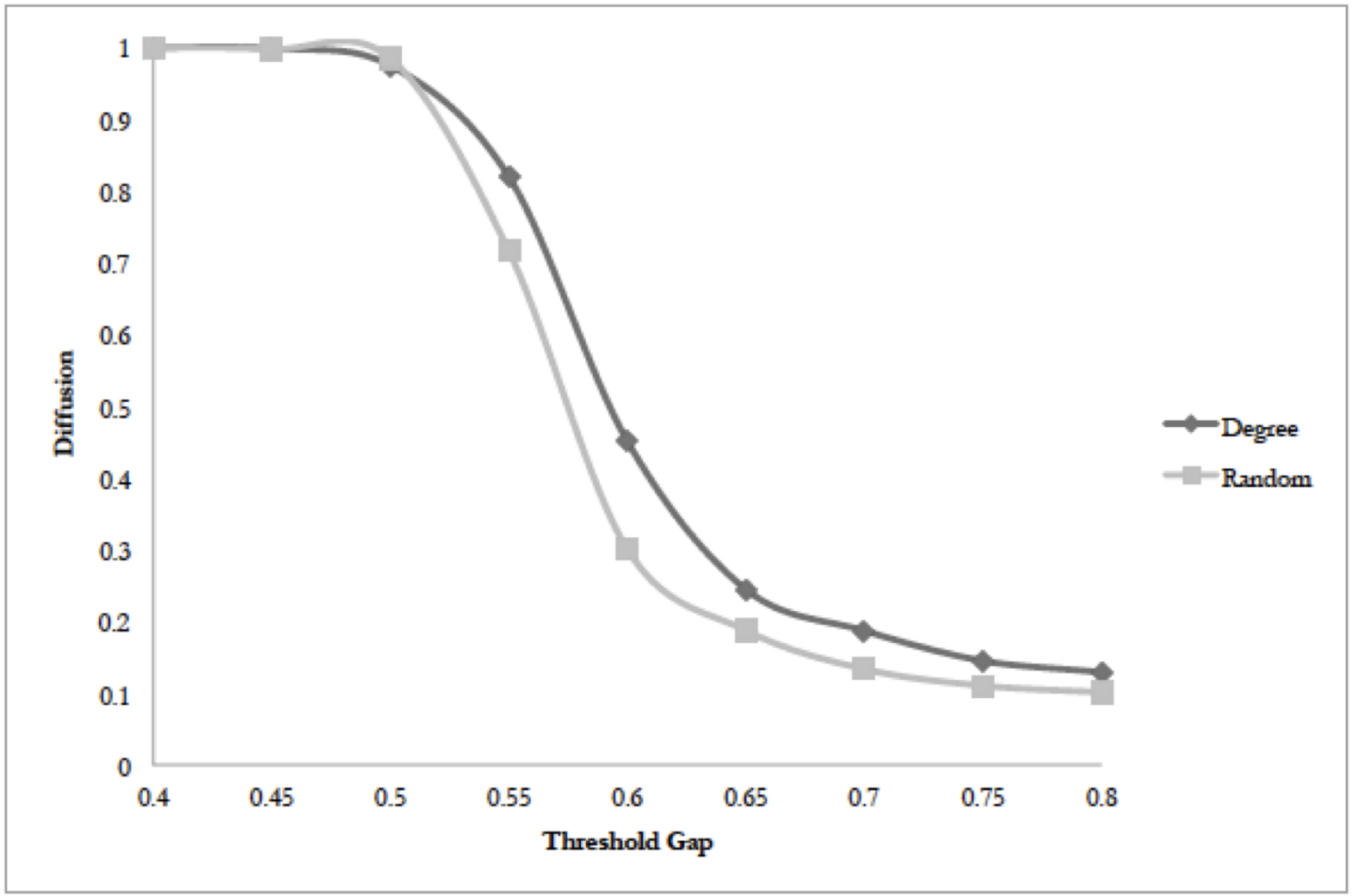
Random Networks



Scale-free Networks



Small-World Networks



Runtimes

Group 1 Heuristics	Runtime (s)	Group 2 Heuristics	Runtime (s)	Random Heuristic	Runtime (s)
D	0.0005	Dw1S	0.0064	R	0.0002
S	0.0052	Sw1S	0.0102		
TS	0.3330	TSw1S	0.3363		
AT	0.0048	ATw1S	0.0105		
ATwSD	0.0053	ATwSDw1S	0.0108		
ATw5Gr	0.0049	ATw5Grw1S	0.0110		
ATw5GrwSD	0.0054	ATw5GrwSDw1S	0.0110		
DD	0.1009				
B	8.4030				
C	17.0212				
E	0.7542				
PR	0.9126				