Graph Analytics class, a.y. 17/18

Third assignment
Social Contagion

0. Introduction

In this assignment, we inspect the diffusion of social contagion.

We use different graphs: the Davis southern women graph, different kind of random graphs (varying the probability of having an edge between nodes), a scale-free graph and a real graph (arXiv collaboration on general relativity).

We try to spread the infection both from a random starting point and from high centrality nodes, using different indexes, in order to see the differences and make observations about the behaviour.

Cantarini Giorgio [s3828113] Franco Danilo [s3809721]

We define also different payoff matrices in order to boost the contagion and evaluate how this choice affects the diffusion on different graphs: the presence of clusters and tightly-knit community could stop the contagion and, in order to reach a final state in which all individual is infected, we propose different setting with increasing contagion rates.

For each graph, we made different plots, one for each contagion starting point (random node, the node with the highest closeness, PageRank,..); then in each of these, we plot different lines, one for each payoff matrix; eventually, these plots help us to see how much time is required to infect all individuals.

1. Background

Epidemic modelling assumes two fundamental hypotheses: *Compartmentalization* and *Homogeneous Mixing*; in this lab, our individuals are in *Susceptible* state or in *Infectious* state and anyone cannot infect anyone else (we need to know the contact network).

We develop the SI Model in which at the beginning all the individuals are susceptible and no one is infected; we assume that an individual (chosen following different settings) become a spreader and the disease will be transmitted in the neighbourhood.

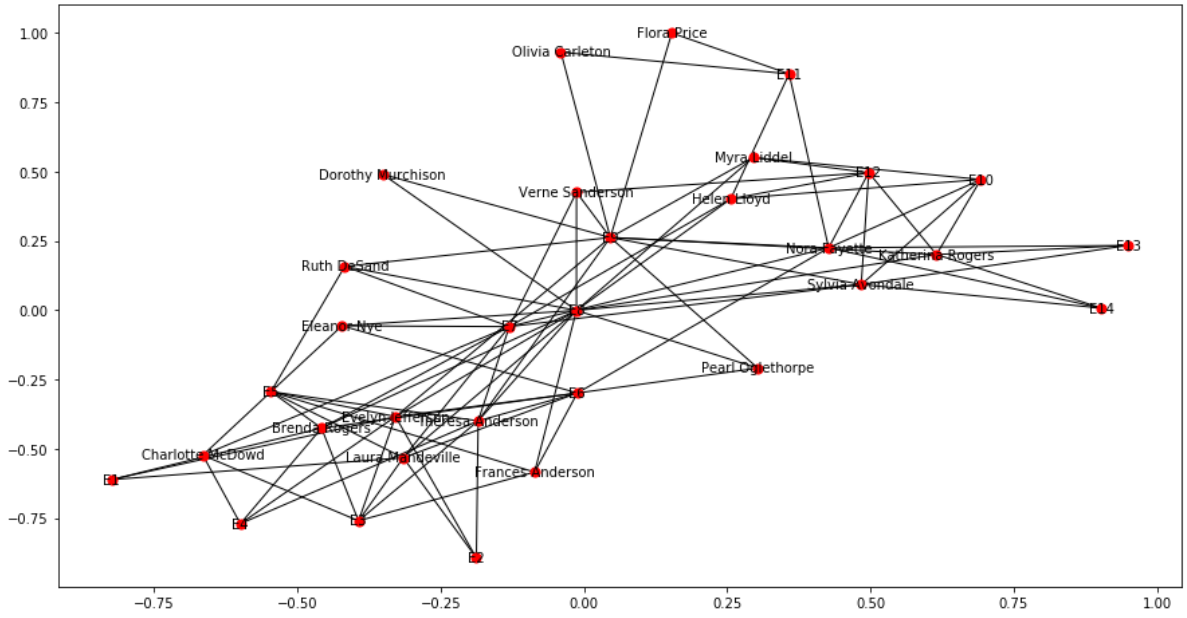
Eventually, the epidemics will end when everyone has been infected.

Following this implementation, there are cases in which the final statement (network completely contaminated) is never reached, so when the contagion reaches a stable state, we infect another node and repeat the procedure until all the individuals are infected.

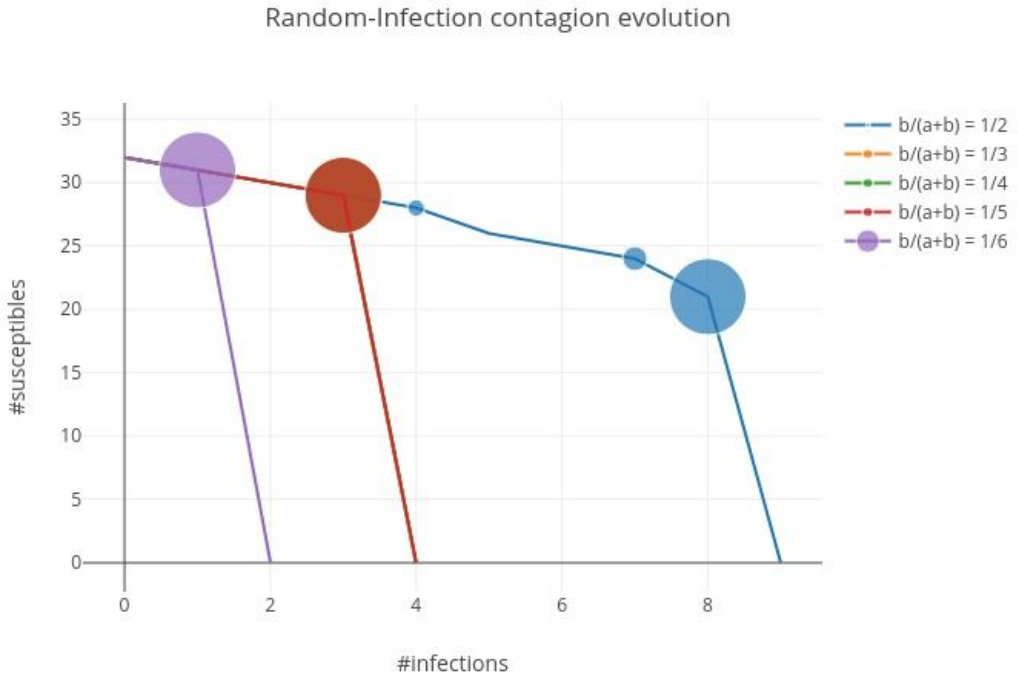
Summing up, the implemented contagion follows a complex system: the chance of having a proper diffusion depends on a number of factors, such as topology of the network, ratio between susceptible and infected nodes and, mainly, on a value that describe how likely is the contagion to spread (e.g. the higher is the fraction of a person's friends that have a good phone, the more likely it is that he would buy one); the main idea is that individuals make decisions based on the choices of their neighbours and on the convenience of changing their behaviour.

2. Experiments

2.1. Davis southern women graph (toy graph)

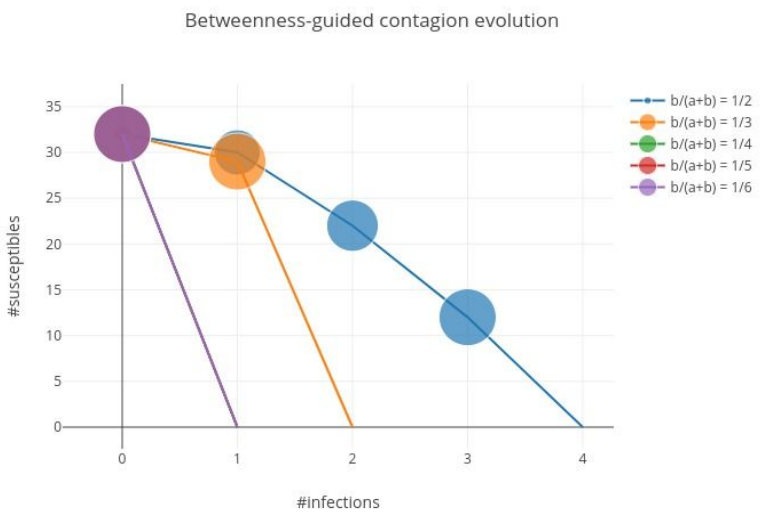


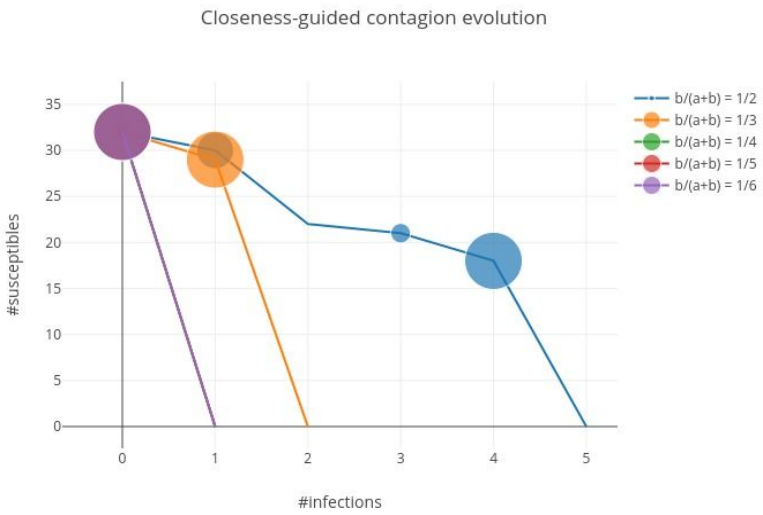
2.1.1. Random infection:

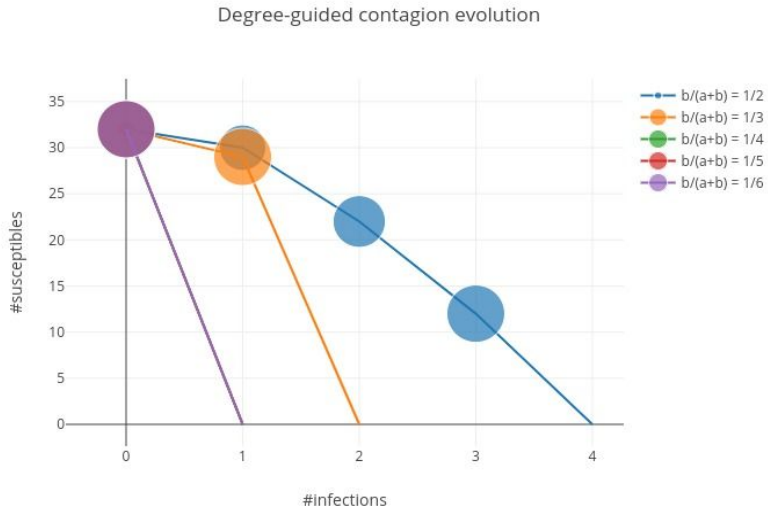


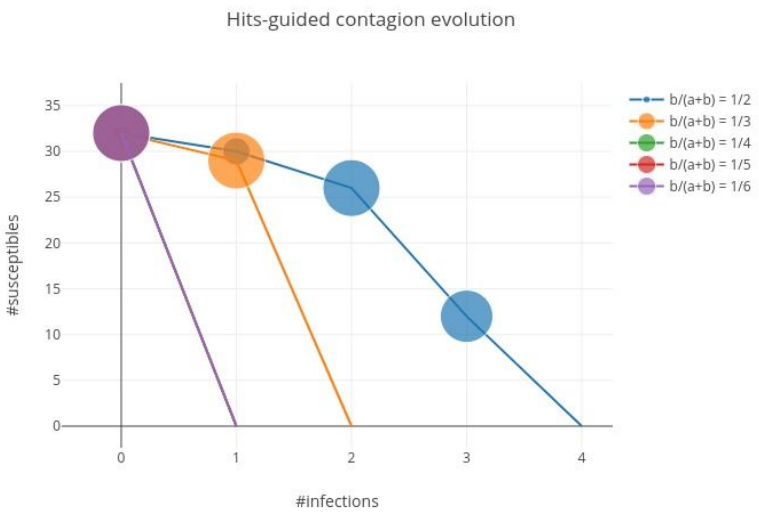
From the plot is clear that using a higher matrix payoff would help the contagion to spread, consequently requiring less time to reach the final state. In this case, using the equal chance pay-off matrix the end state is naturally never reached (requiring repeated infections).

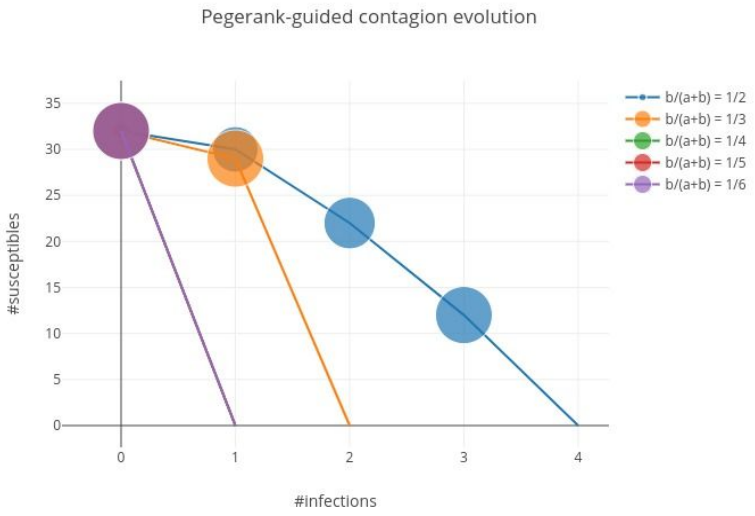
2.1.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:











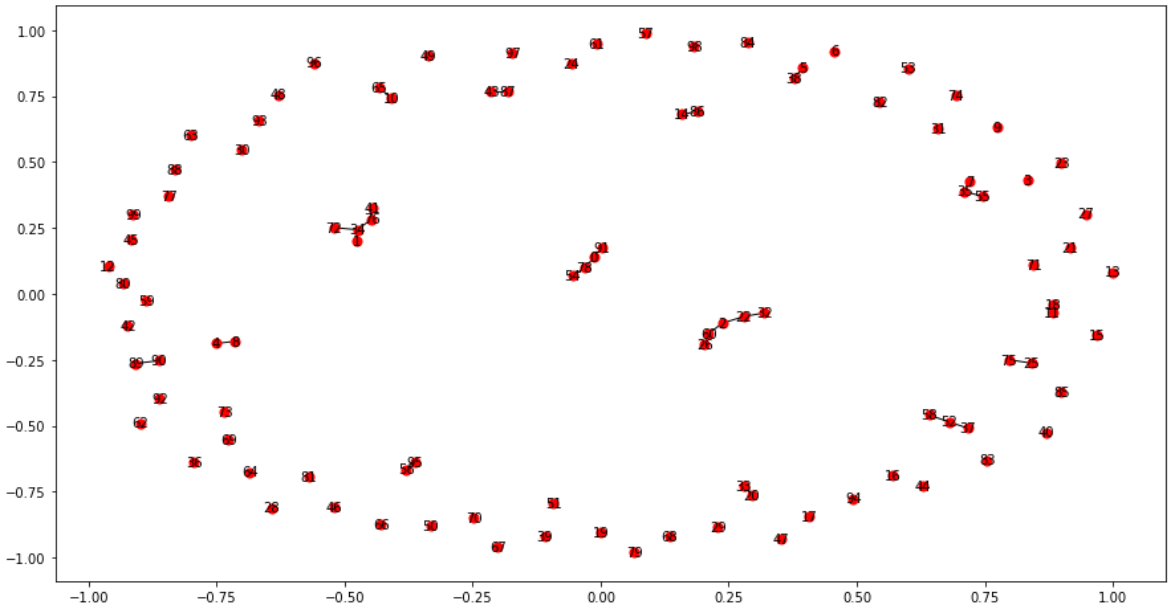
From these plots, it's obvious that the contagion will require less time to spread to the whole graph: selecting the most important nodes (according to some centrality index) we could infect more neighbours in the first and subsequent steps, with lower structural impedance in the early stages.

Using the equal chance pay off matrix the best results in terms of time (or the number of infections in our model) are reached starting the contagion from the nodes with the highest clustering coefficients (spreading inside cliques).

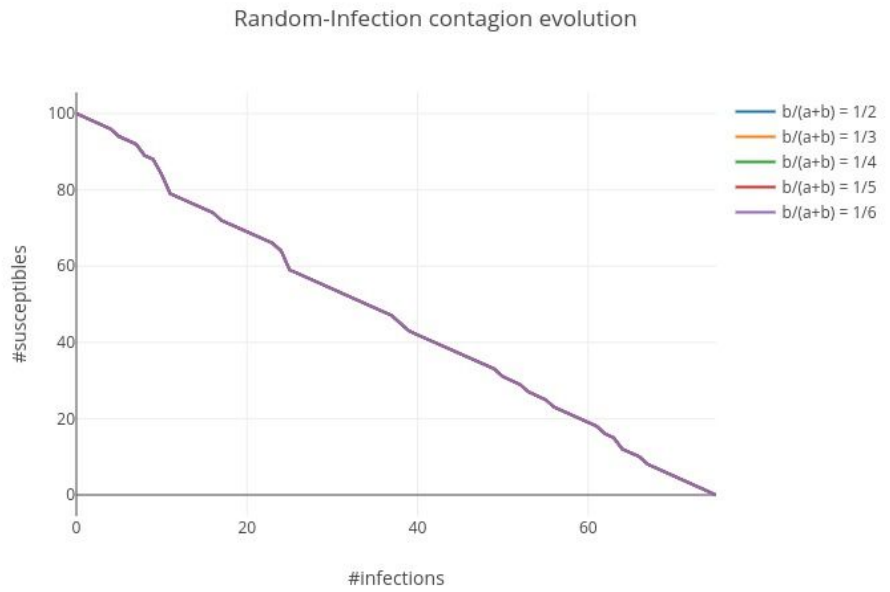
2.2. Random Graphs

2.2.1. Subcritical regime

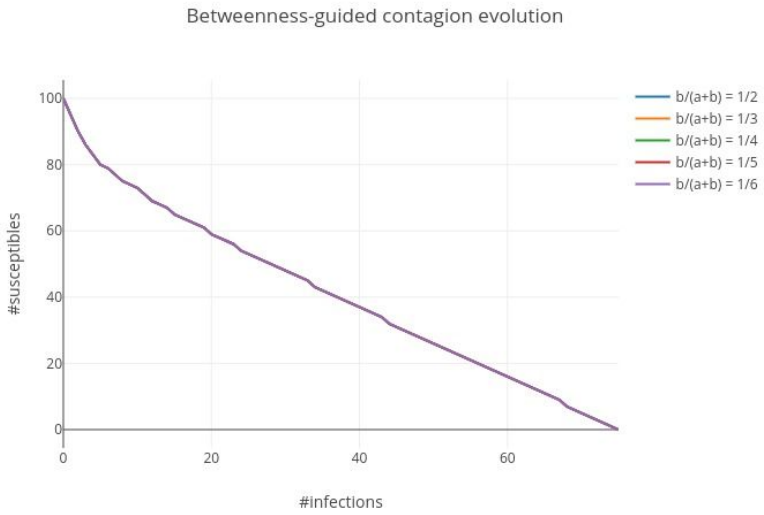
$$\left[p < \frac{1}{n} \Rightarrow 0 < \langle k \rangle < 1 \mid \langle k \rangle = 0.5\right]$$

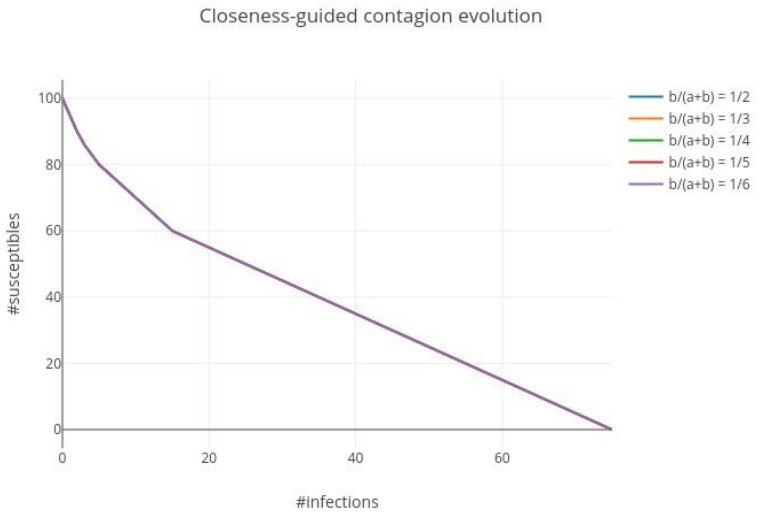


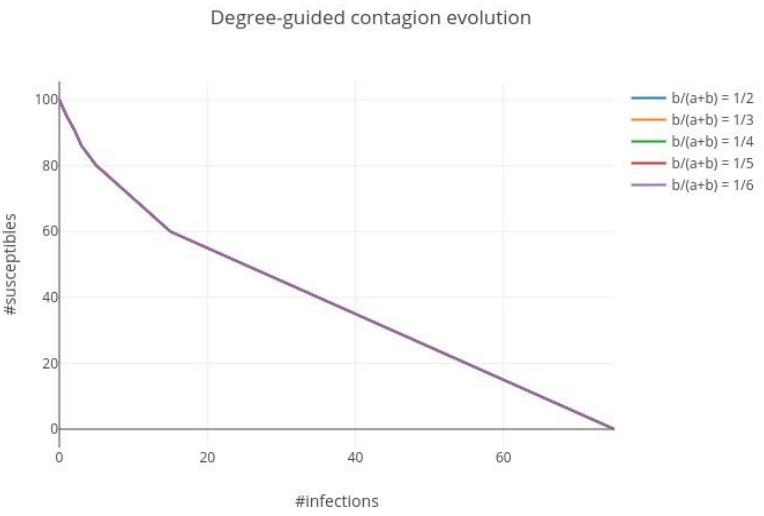
2.2.1.1. Random infection:

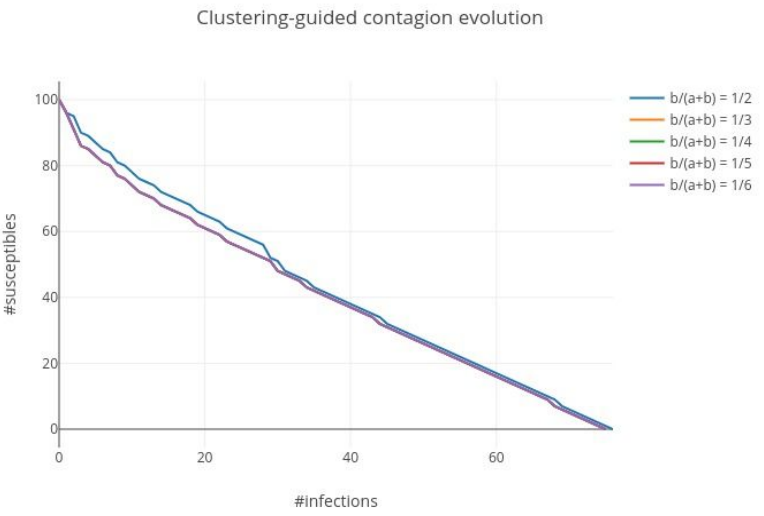


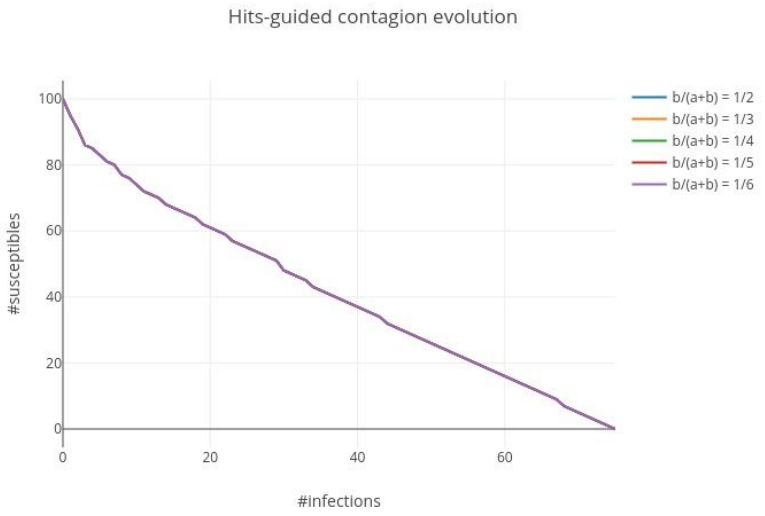
2.2.1.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:

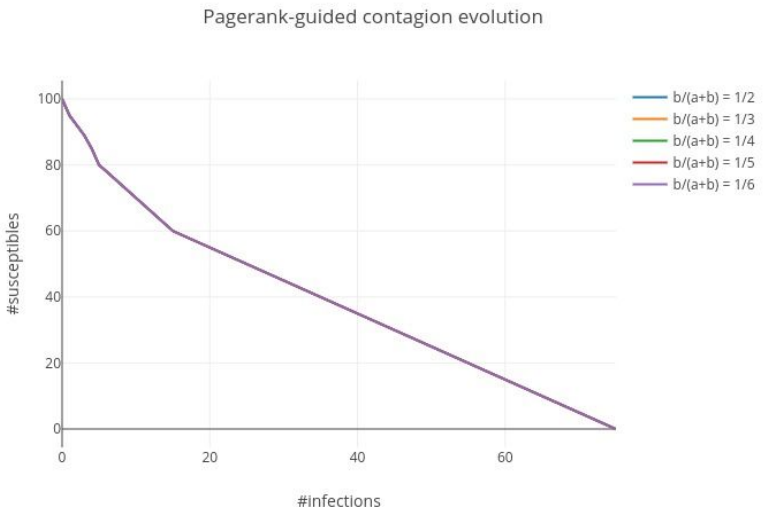








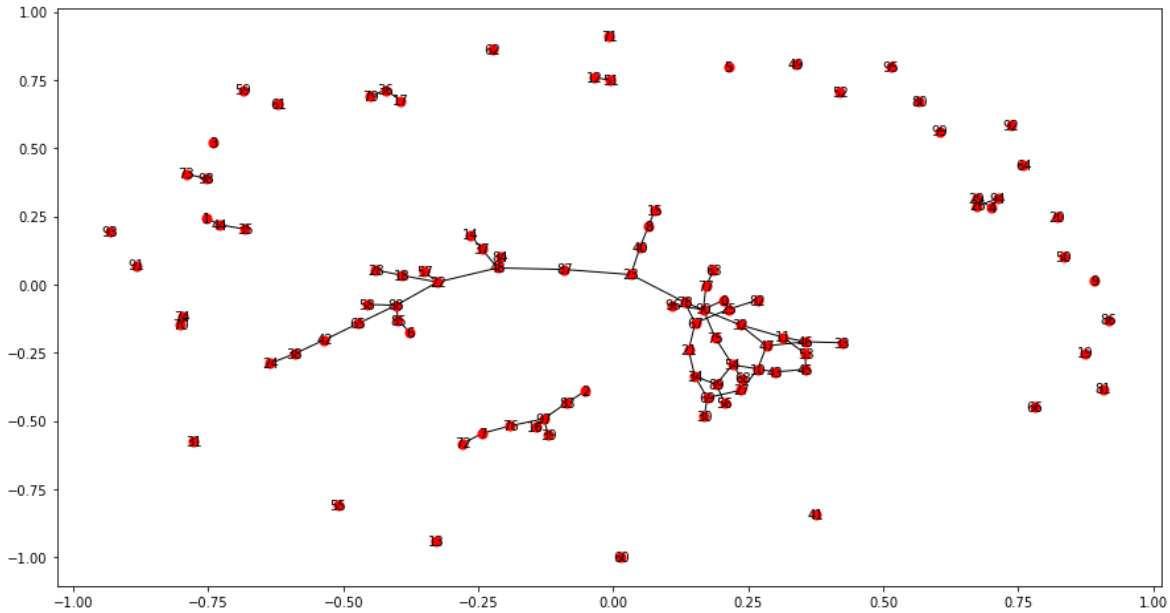




Being composed of different tiny components the spread of infections cannot reach all the nodes: when the contagion stops, we repeatedly infect randomly another node; reaching the final stage requires a lot of time, whatever setting is being adopted (this is due to the architecture of this network).

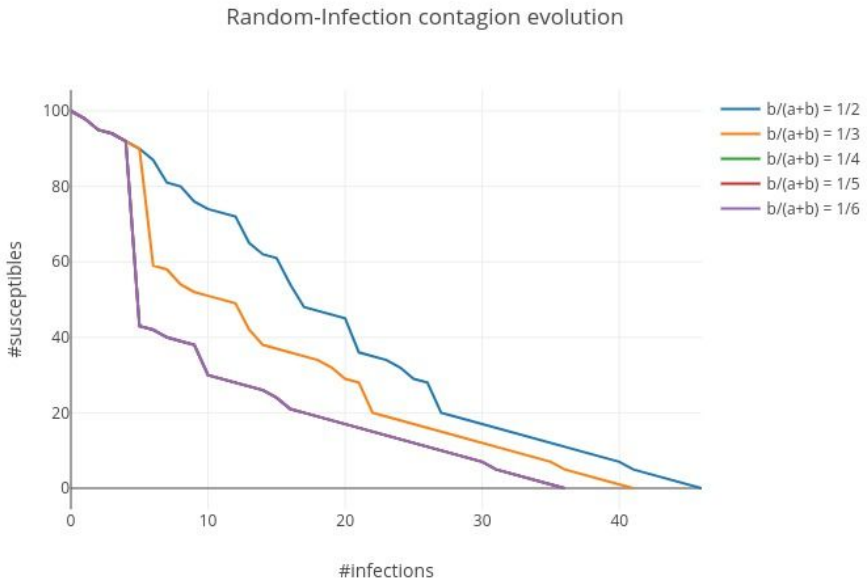
2.2.2. Supercritical regime

$$\left[\frac{1}{n} < p < \frac{\ln(N)}{N} \Rightarrow 1 < \langle k \rangle < \ln(N) \mid \langle k \rangle = 1.36\right]$$

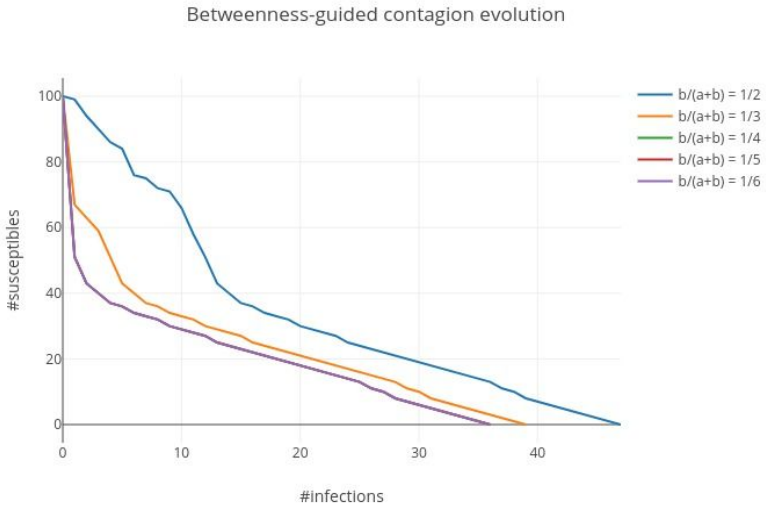


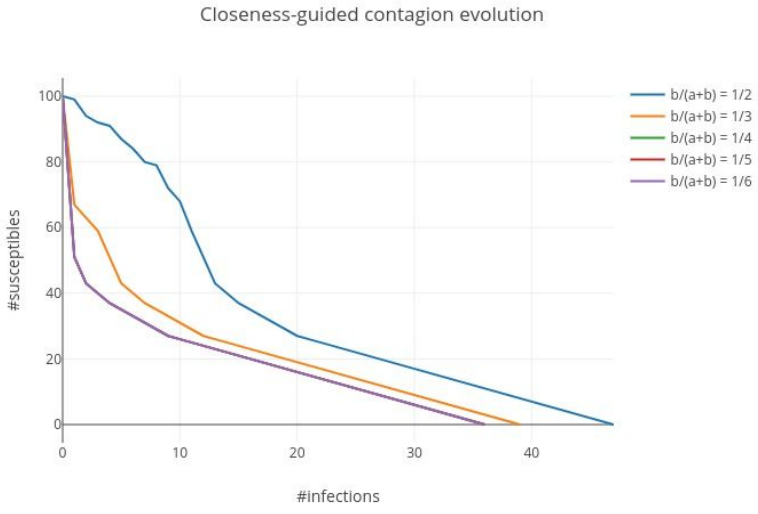
The giant component looks like a network (in the sense that it contains a finite fraction of the nodes): while numerous isolated aggregates are shaped like trees, the giant component contains loops and cycles.

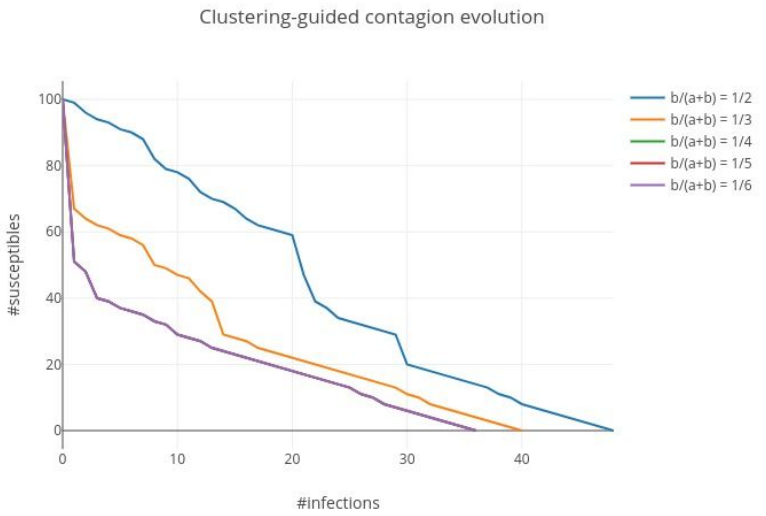
2.2.2.1. Random infection:

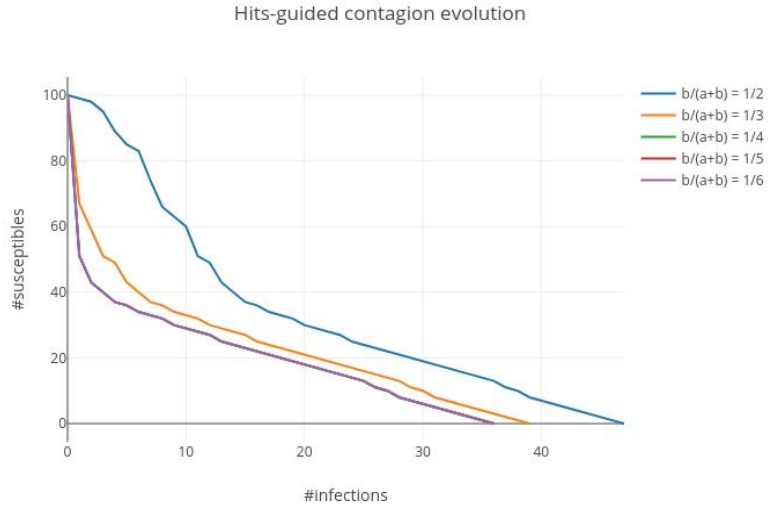


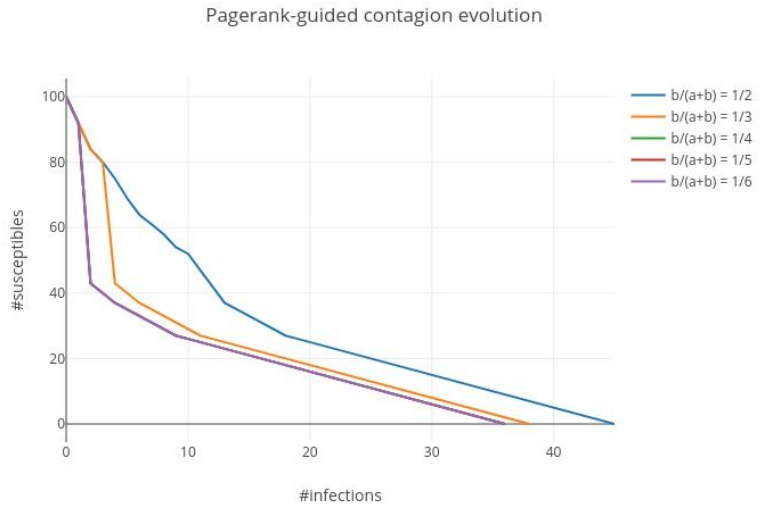
2.2.2.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:







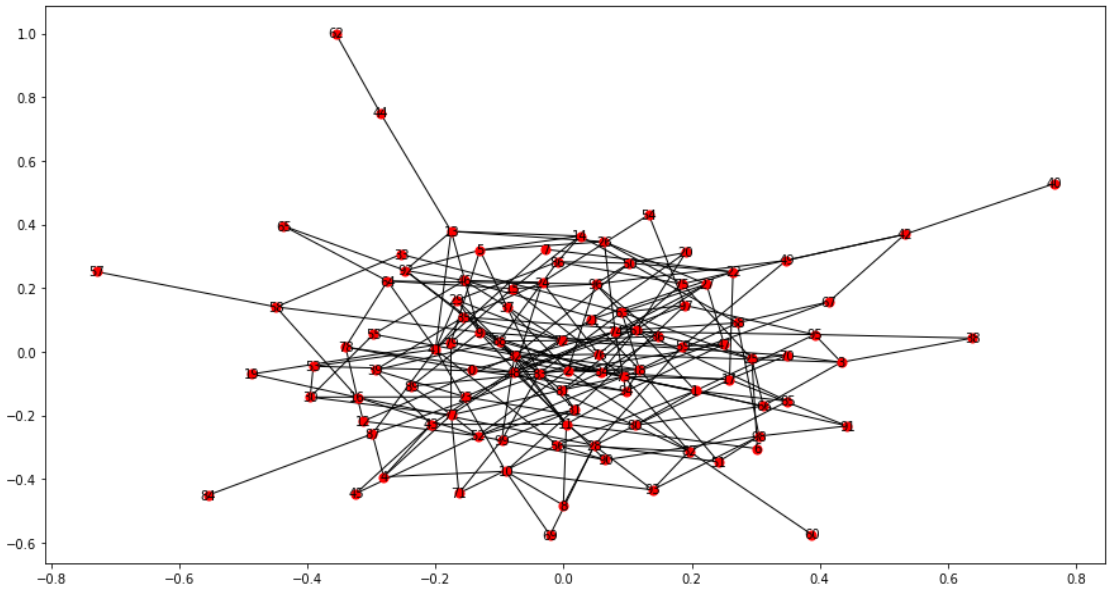




Intuitively, different plots do not change evidently between a random setting and centrality-driven one: it suggests that the graph is still too disconnected and we cannot see relevant spreading behaviour. However, the discrepancy of the equal chance payoff matrix starts to emerge: it requires more time w.r.t the unbalanced matrices (that boost the spreading).

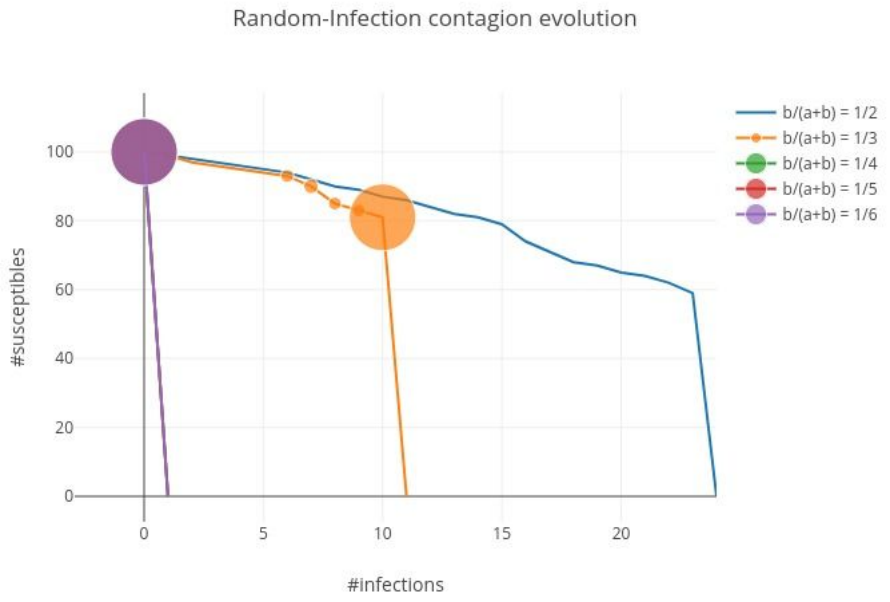
2.2.3. Connected regime

$$\left[p > \frac{\ln(N)}{n} \Rightarrow \langle k \rangle > \ln(N) \mid \langle k \rangle = 5.44, \ln(N) = 4.72\right]$$



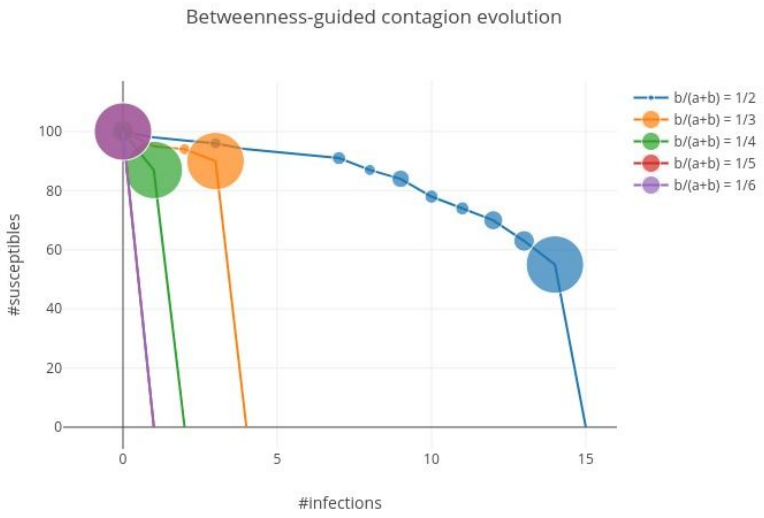
For sufficiently large p the giant component absorbs all the individuals: if isolated nodes are not present, the network is defined as connected.

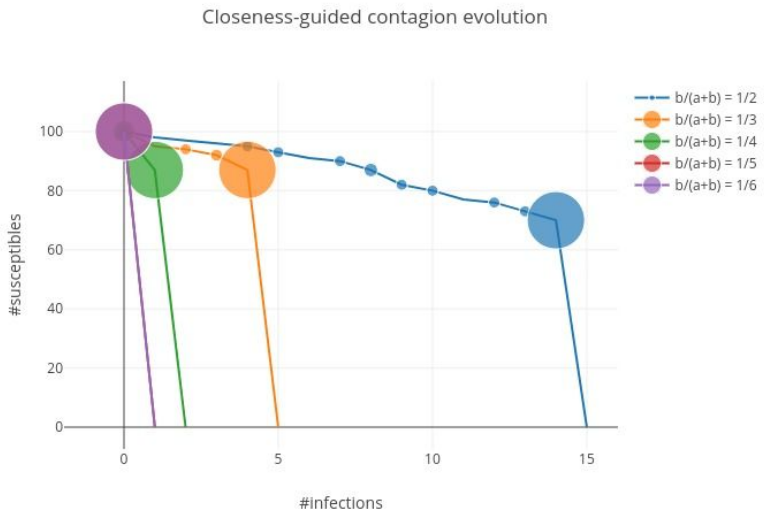
2.2.3.1. Random infection:

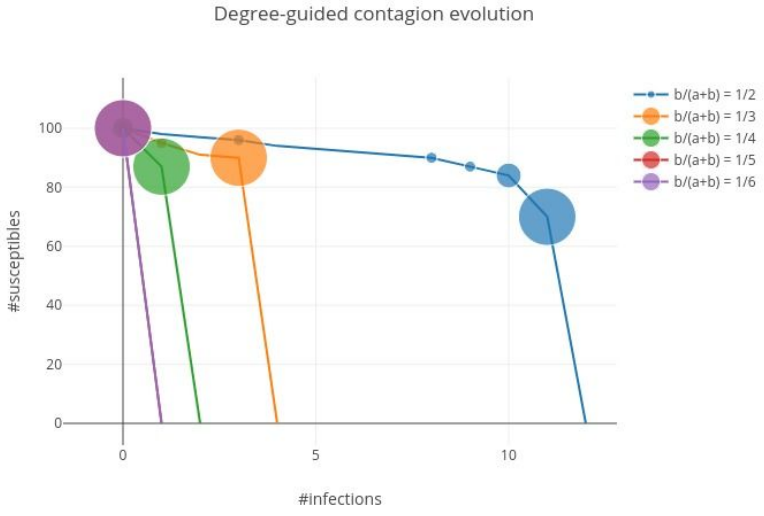


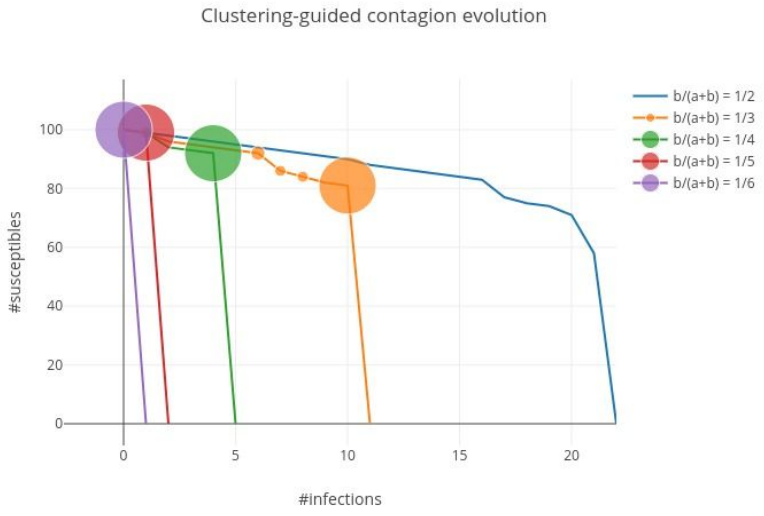
Looking at the plot, we can see easily spot the differences: the equal chance contagion requires less time compared to the other regimes (subcritical and supercritical); similarly, the usage of increasingly unbalanced payoff matrices boosts the spreading rate and reduces the time required to infect everyone.

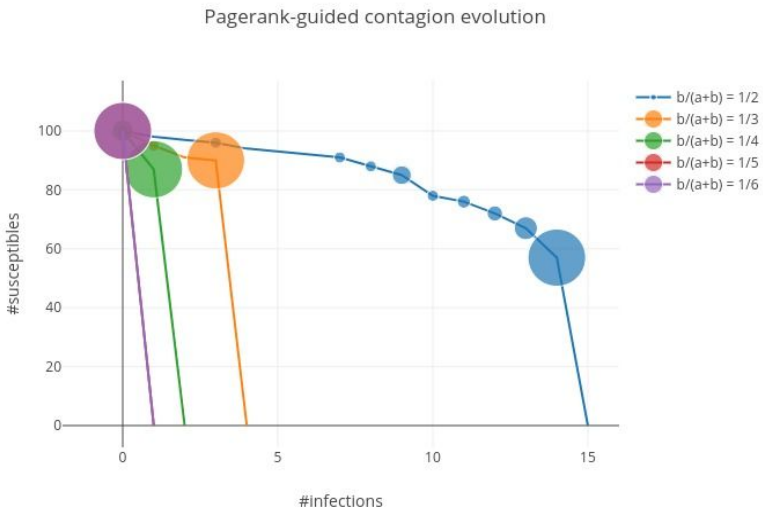
2.2.3.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:

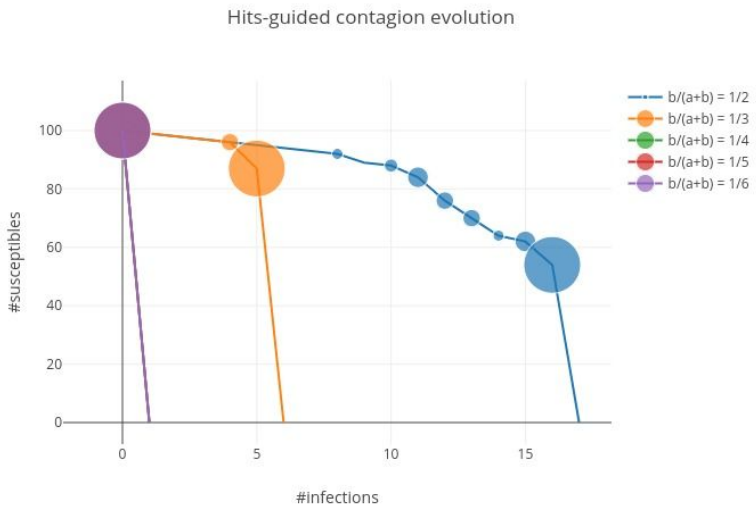






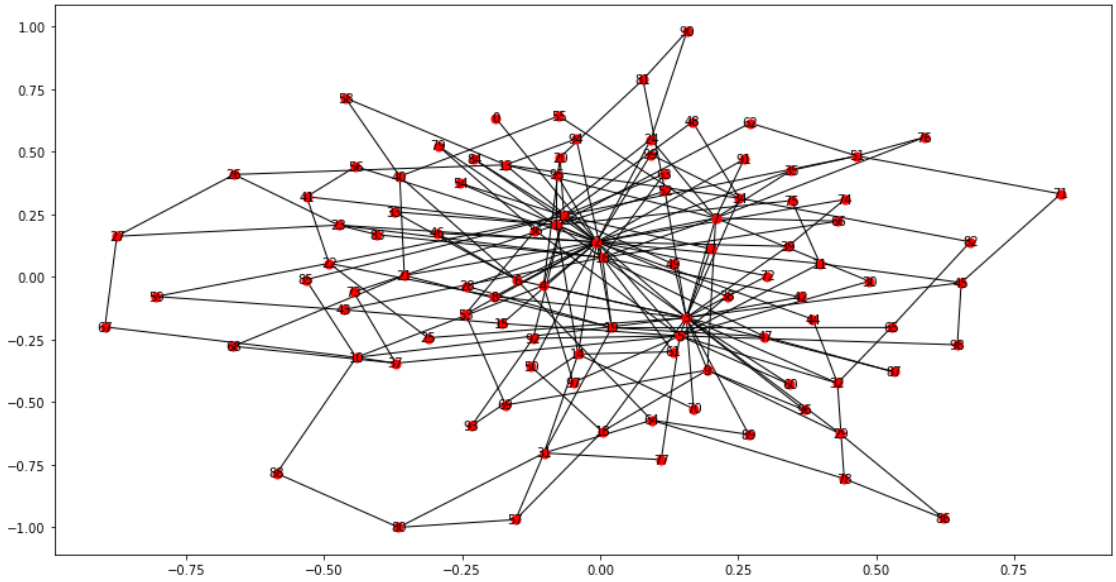




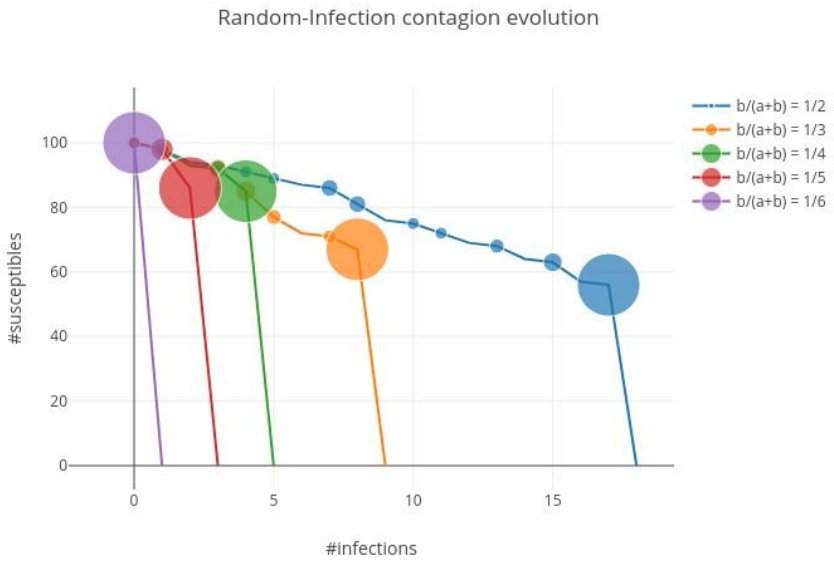


With different choice as a starting point, we achieve better results (as expected): the measure that gives the best result with the equal payoff matrix is the betweenness (or, alternatively, PageRank); more or less, they are all comparable, except for the clustering coefficient index (we will comment in the following paragraphs where this behaviour is even more accentuated).

2.3. Scale-free Graphs



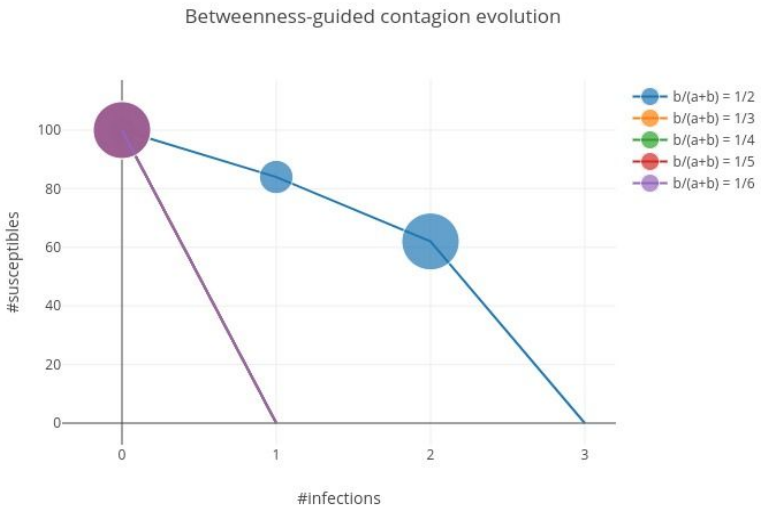
2.3.1. Random infection:

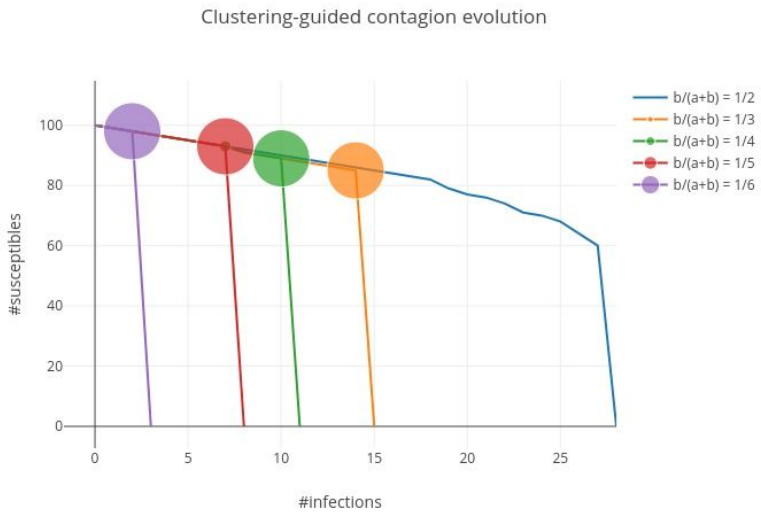


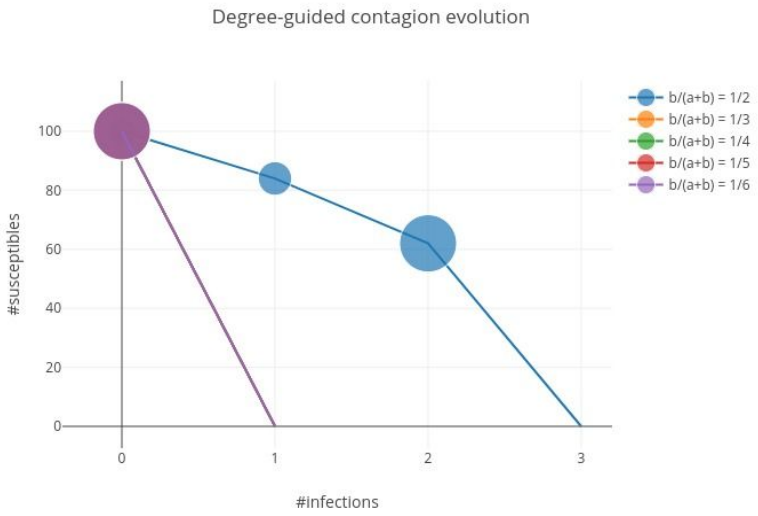
The presence of hubs noticeably slows the pace of random infection: intuitively, the node majority possesses a low amount of connection, the random contagion is consequently very likely to pick a low-spread individual; unfair matrices, however, mitigate this natural impedance.

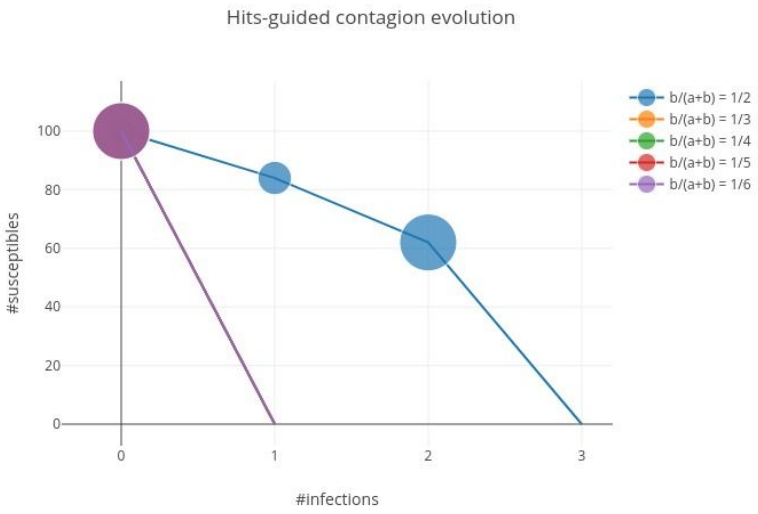
2.3.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:

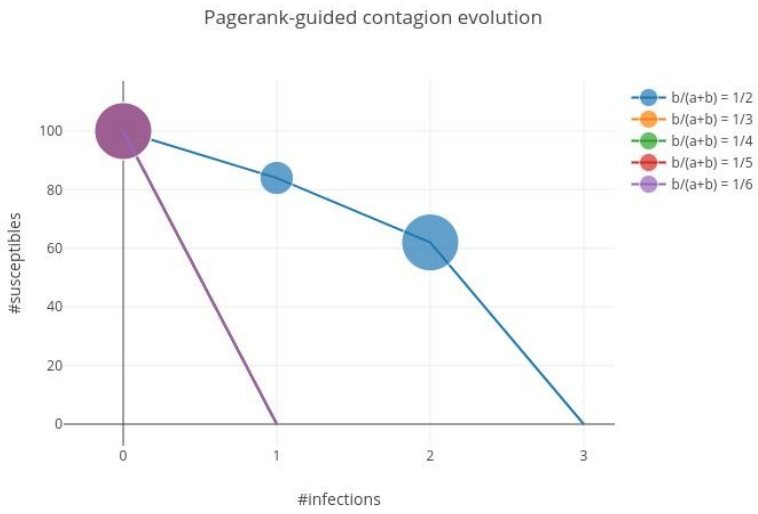
In this setting, the centrality-driven infection makes possible the phenomenon of super-spreaders: being able to affect immediately particular nodes (i.e. the hubs), few infection instances are needed in order to reach the epidemic stage.







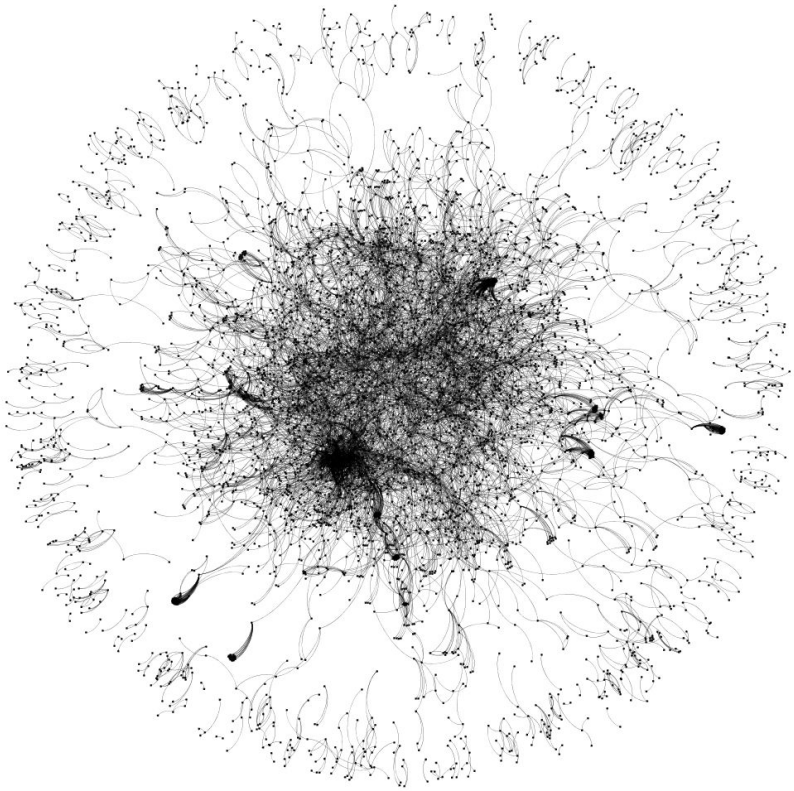




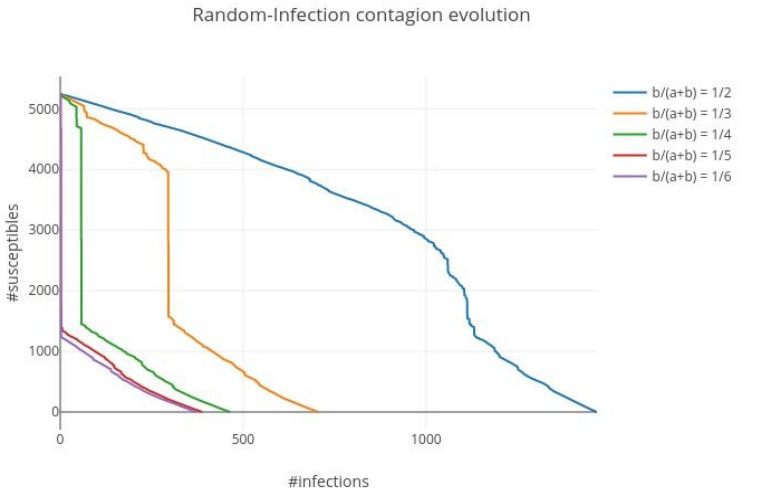
The only measure that differs much from the others is clustering: this is due to the fact that this factor is related to the number of connected nodes as a clique (increasing N the time required to spread the infection using clustering coefficient as a centrality measure grows).

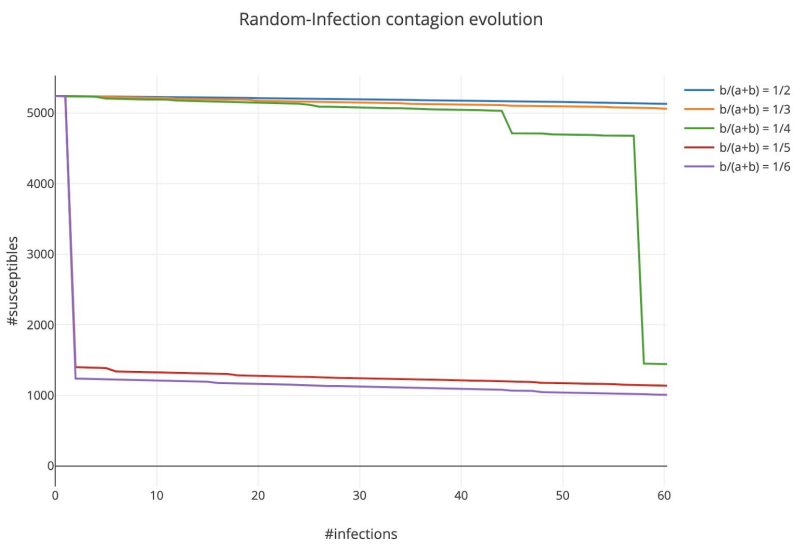
If we pause for a moment and think about the clustering coefficient driven contagion, it means that we're trying to infect a highly connected component: in a clique, the susceptible nodes (not infected yet) act strongly against the spreading rate so we will not see a break-out until the majority of nodes are infected, from which point, the entire connected component will act as a single hubs.

2.4. Real graph (Arxiv GR-QC)



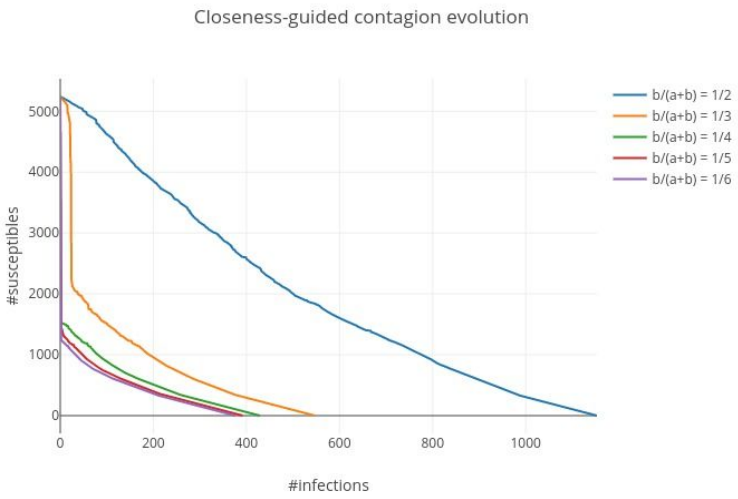
2.4.1. Random infection:

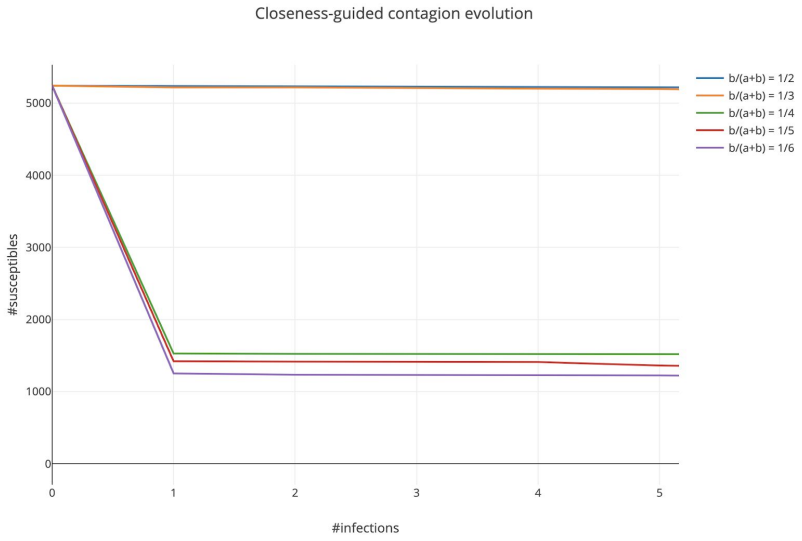


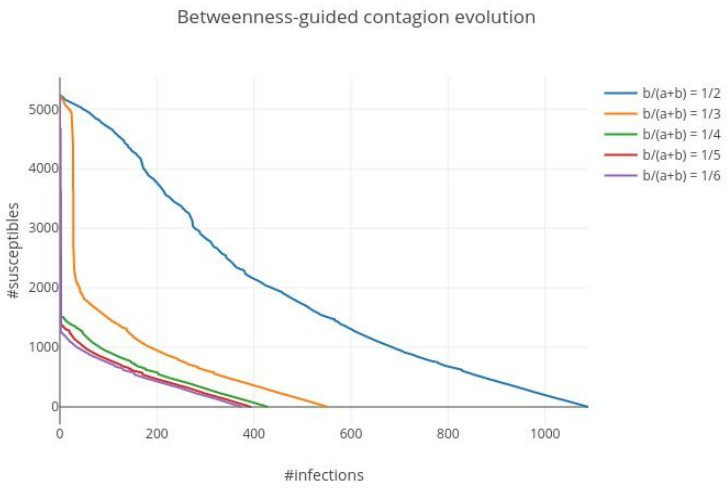


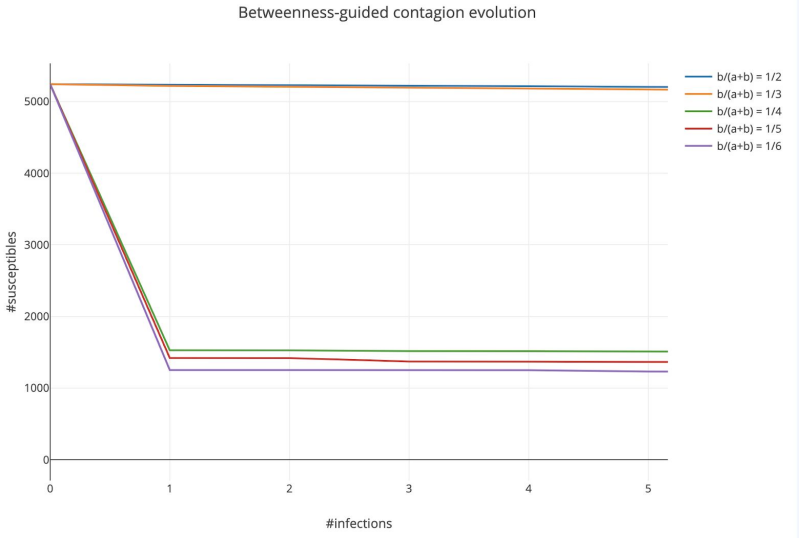
For every pay-off setting, we can observe the same pattern: a first stage where the contagion starts to pick random individuals who alone can't start a contagion burst, thus decreasing linearly the number of susceptibles; then, when a sufficient number of nodes is infected (determined by the spreading rate), an exponential breakout definitely happens: at this point, the entire giant component is being infected (we observe, in fact, that the exponential decrease stops more or less at the same quantity, representing the individuals living in the main connected component); finally, pretty much the same linear decrease is carried out(in term of steepness), meaning that we are left with the contagion in the sparse periphery.

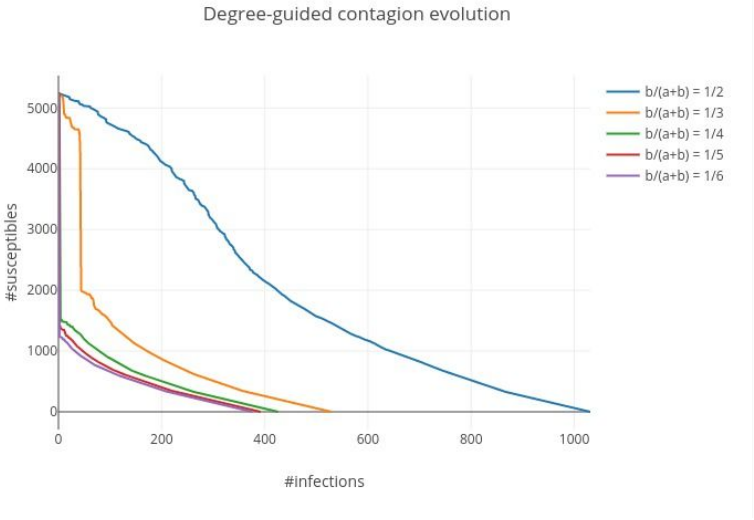
2.4.2. Betweenness, Closeness, Degree, Clustering, PageRank, HITS Infections:

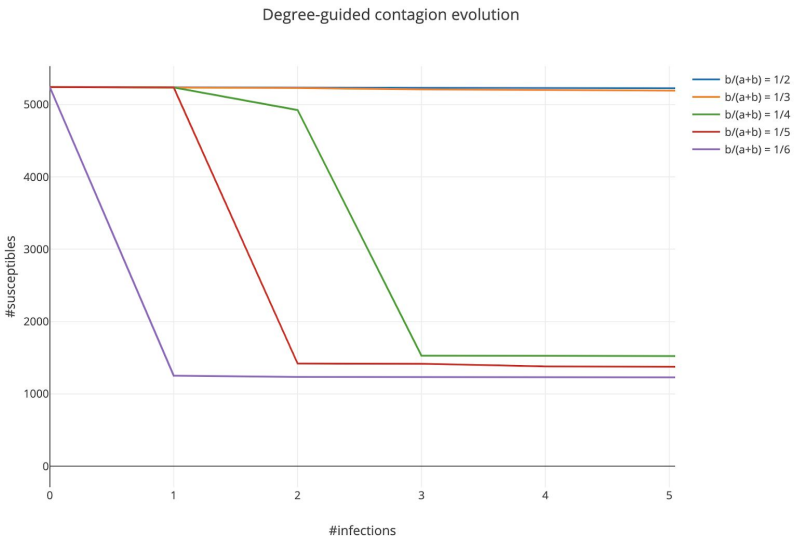


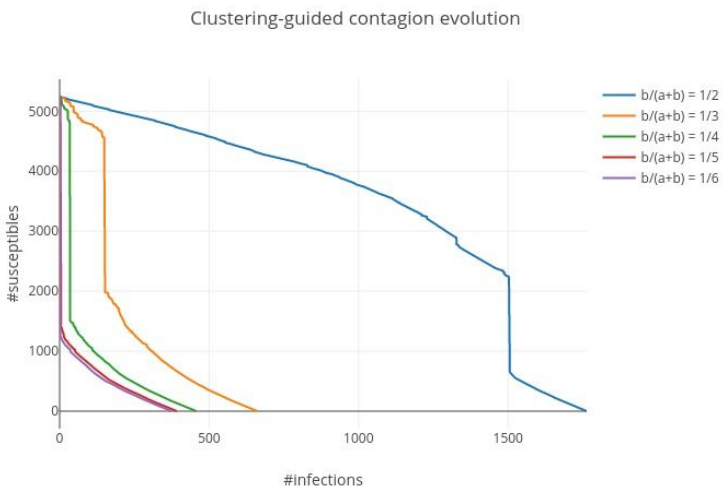


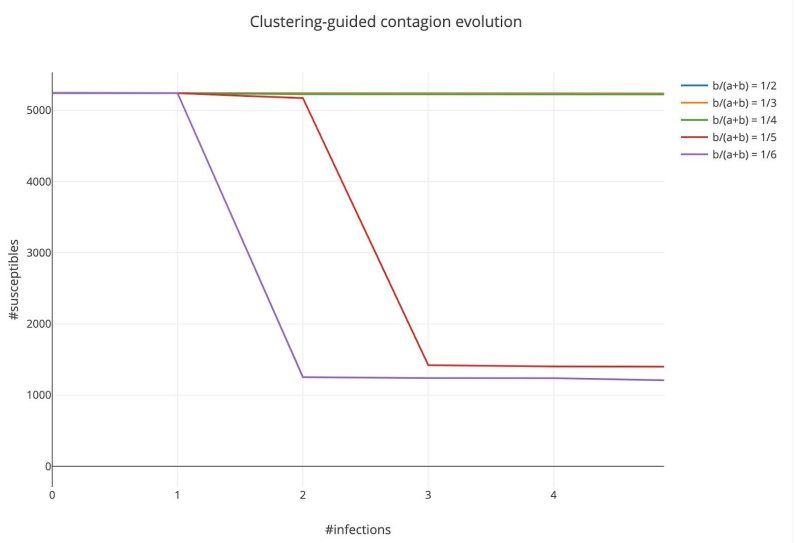


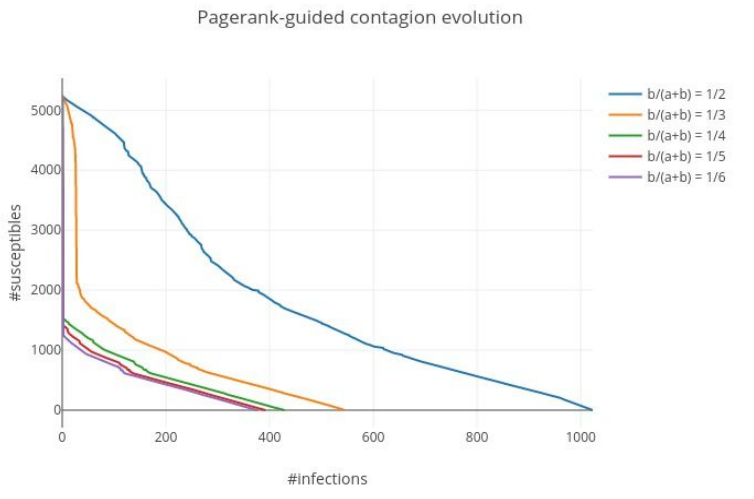


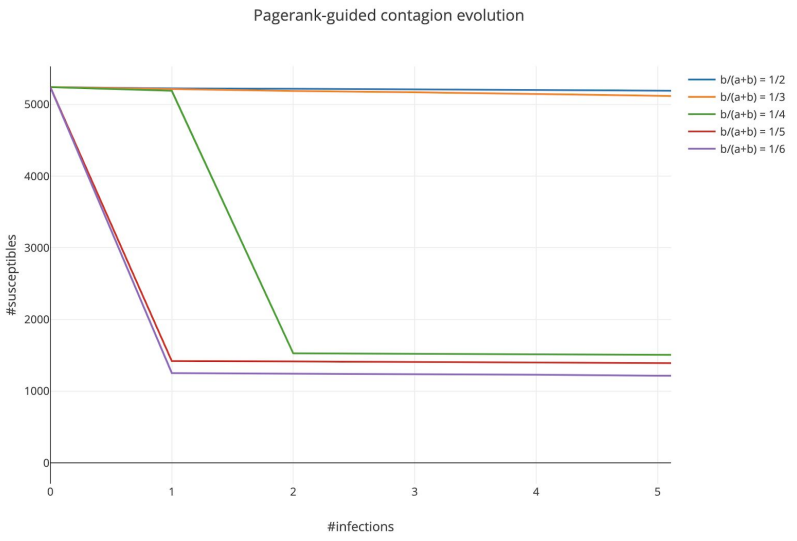

Degree-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections

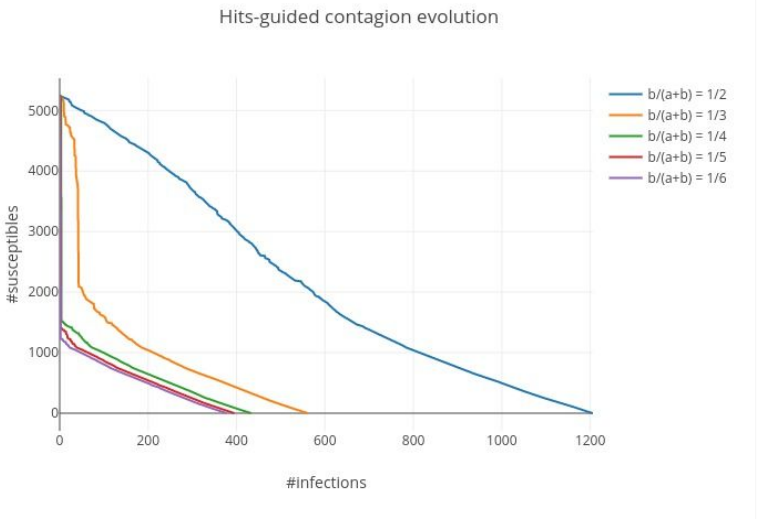

Degree-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections

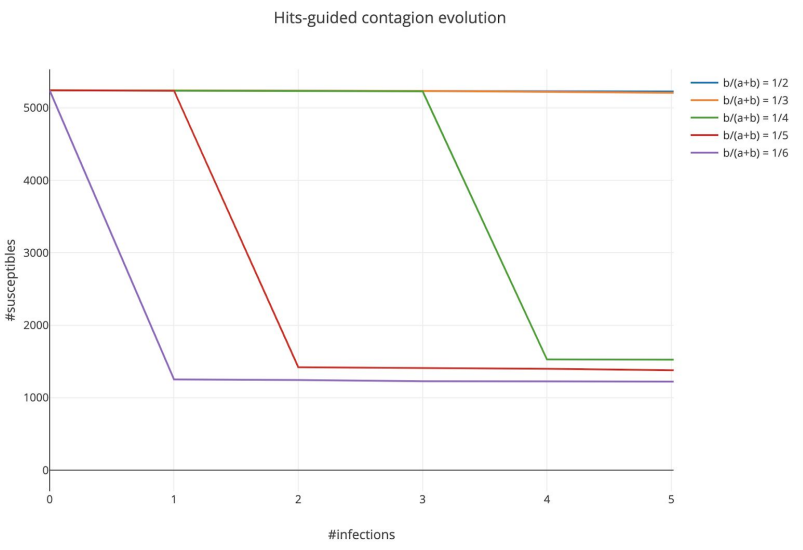

Clustering-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections


Clustering-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections


Pagerank-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections


Pagerank-guided contagion evolution
b/(a+b) = 1/2
b/(a+b) = 1/3
b/(a+b) = 1/4
b/(a+b) = 1/5
b/(a+b) = 1/6
#susceptibles
#infections





The centrality-driven contagion shows similar behaviour to the one viewed for the random approach, effectively the three phases are easily spottable in most cases; the best results in term of speed are obtained for the betweenness and closeness centrality: for them, the ¾ payoff infection obtains the most rapid breakout, followed by the PageRank index.