Complex Systems and Network Science Project Report

Giovanni Pio Delvecchio

giovanni.delvecchio2@studio.unibo.it
Student ID: 0001037185
Alma Mater Studiorum University of Bologna

Abstract. In the paper published by Scott L. Feld on the American journal of sociology, Why Your Friends Have More Friends than You Do [1], an analysis of a significant sociological "class size paradox"[2] is addressed, which is: Individuals find themselves associated with people with more friends than they have, thus they could conclude that are below average and inadequate. While Scott L. Feld performs a detailed and mathematical analysis of this phenomenon using real word data, in this paper it will be performed an analysis about friends in networks, mainly about the distribution of friends (neighbors), the distribution of friends of friends and the distribution of the mean value of friends of friends in the context of two different classical network configuration models (Erdős-Rényi [1] and Barabási-Albert [2]). Moreover, the statistics about friends will be computed in the context of different types of network dynamics: random, based on a genetic algorithm and based on a variation of the Treaty Game described in [3]. Further analysis will be carried out in order to better understand the characteristics of the involved networks.

Keywords: Networks · Network Science · Network Dynamics · Social Networks

1   Introduction

Analyzing the relationships between the number of friends and the mean number of friends of friends of an individual is the main topic of the paper by Scott L. Feld [1]. However the focus of [1] is on real world data, from which is possible to draw conclusions only about a "snapshot" of real-word friendships.
The objective of this paper is to analyze friendship networks in a dynamic setting, where individuals can become friends with other individuals or interrupt the friendship relationship by means of random behaviour, similarity between the characteristics of each individual or the improvement of a payoff score.
In order to perform this kind of analysis, various experiments must be carried out, hence it is necessary to start from known baselines in order to obtain results

---

[1] Erdős-Rényi model: https://en.wikipedia.org/wiki/Erd%C5%91s%E2%80%93R%C3%A9nyi_model
[2] Barabási-Albert model: https://en.wikipedia.org/wiki/Barab%C3%A1si%E2%80%93Albert_model

that are not strictly related to a single real-word setting, even because the latter takes into account a plethora of factors, like shared interests and social norms, that would require an ad-hoc modeling in a dynamic environment.

Classical models for networks are taken into account because of their limited number of parameters, which can be tweaked in order to make the dynamics more predictable, but the use of more sophisticated models is possible nonetheless.

2   Description of the baseline models

In order to analyze the statistics about friends without using real-word data, it is possible to choose one of the most well-known network configuration models, two of which have been chosen for this study and are briefly described below (note that since the objective is to model friendships it is not necessary to consider directed graphs):

- The Erdős-Rényi model, where n nodes are generated and each of the $\frac{n(n-1)}{2}$ possible edges is generated with probability `edge-probability`. It is unlikely that friendships will form in such a way, since for high values of `edge-probability` we obtain a highly connected graph. It is possible to use this model as a starting point for the network dynamics methods here presented nonetheless, since for low values of `edge-probability` we obtain graphs with similar values of average number of friends and average mean number of friends of friends (generally low variance, where variance is the property described in the section "Variance affects the mean number of friends of friends" of [1]).
- The Barabási–Albert model, where starting from 2 connected nodes we add $n-2$ nodes using preferential attachment, meaning that each node `new-node` will link to one of the existing nodes `old-node` with probability equal to

$$\frac{\texttt{degree(old-node)}}{\texttt{sum of the degrees of all nodes}}$$

This is a good model for environments where friendships are formed, if we consider that the dynamics of "Rich-get-richer" is reflected in the context of popularity (those that are already popular tend to be more popular by meeting other people). However this model is the one with the highest variance (in general > 1) with respect to the others here presented.

3   Description of the models for the dynamics of the networks

Here are presented three models for the network dynamics, in ascending order of complexity. A detailed description of how the parameters were chosen can be found in the next section. The models are:

– A simple random dynamics algorithm based on the algorithm associated to the Watts-Strogatz model: considering two nodes `a` and `b` it rewires one of the edges of `a` (if it exists) to have `b` as target node, whith probability `rewire-probability` (the single parameter of this algorithm).
  With respect to the dedicated approach for the Watts-Strogatz model, we have two main differences: it does not necessarily need a regular ring lattice as a starting baseline and it is possible to add a link between a node that has no edges and another one (chosen randomly).

– A genetic algorithm for network dynamics that works as follows:
  1. A chromosome (essentially a randomly generated 0-1 vector with dimension `vec-size`) is assigned to each node.
  2. At each time-step we consider the `num-to-replace` nodes that have the smallest degree and change their chromosomes, then for each chromosome of the population we mutate a gene (flip a bit) with probability `mutation-probability`.
  3. For each node `a` we consider the cosine similarity between his chromosome and all the other chromosomes of the population, this similarity is then re-scaled according to its degree and the one of the other nodes, so that for each other node `b` we construct the following vector:
  $$similarities(\mathtt{a}) = [cosine - similarity(\mathtt{a}, \mathtt{b}) * (\tfrac{degree(\mathtt{a})}{degree(\mathtt{b})})$$
  $$\forall \mathtt{b} \in N]$$
  where N is the set of all nodes in the graph. An edge is created between `a` and the first `num-new-links` entries of this vector, after that we drop the links of `a` with probability `drop-links-prob` in order to add noise and avoid to create highly connected networks.
  4. if all the nodes have at least `min-deg-to-reach` edges, we stop.

  In this algorithm the chromosomes play an important role: they represent the characteristics of each individual and define if it is similar or not with respect to another one, conditioning the decision of a node to connect to another node with similar degree (a node would chose to connect to another node with similar or lower degree, prioritizing those with the lowest degrees, so the fraction between the degrees represent the fitness function of each node, in fact we change those nodes that "do not fit well" during step 2).

– A game theory algorithm for network dynamics based on the Treaty Game described in [3] that works as follows:
  1. Each node decide to sign or abstain from signing a treaty randomly with probability 0.5 (arbitrary value). Nodes that decide to collaborate will do so by contribute a with a fixed amount in the range [`y-max`/2, `y-max`], those who abstain will contribute with amount in range `0, y-max` or be selfish and contribute 0, again with probability 0.5 (this value is arbitrary too).
  2. the payoff for each node that contributes is computed as:
  $$\pi_i = (y_{max} - y_i) + b_{contribute} * (y_i + y_{-i})$$
  while the payoff for each note that abstain is:
  $$\pi_i = (y_{max} - y_i) + b_{abstain} * (y_i + y_{-i})$$

Where y_i is the contribution of the node i and y_{-i} is the sum of the contribution of the neighbors of the node i. The two marginal gains $b_{abstain}$ and $b_{contribute}$ are a simple mechanism for punishing nodes that do not collaborate from the start and reward those that did.

3. at each time-step the `n-nodes-gossiping` nodes with the smallest payoff are involved in a gossip procedure where they share the information about their contributions and the view of the contribution of their neighbors and decide whether to create an edge (or destroy it) with a node of the view based on the gain that creating the edge they would have on the payoff (the gain is expressed as the percentage `gain-percentage`, for which the choice is 110%, which empirically works well in this setting). After changing the edges by this means, they are dropped with probability `drop-prob`.

4. The value of the contributions are static until nodes have a payoff lower than `payoff-threshold` or a degree lower than `num-nodes-thresh`, in that case the "unsatisfied" nodes will change strategy according to what stated in the first step.

5. If each node has at least `min-deg-to-reach` friends, the algorithm stops.

Similarly to the Treaty Game we have the first three initial steps (collaborate/abstain, fix the contribution for collaborative nodes, let each agent that abstain decide whether it should contribute or not), a maximum contribution value and the "chicken" characteristic (each node would benefit by not-signing the treaty, but cooperation allows the achievement of a better payoff).

However, the decision of the nodes that sign the treaty is to collaborate with more than half of the value (in the Treaty game collaborative nodes would have a contribution in range $[0, \texttt{y-max}]$), and the nodes that abstained can choose to contribute 0 or with a value in $[0, \texttt{y-max}]$ (in the treaty game it would be always 0 because it is just more convenient), this choice is done because we want to use the payoff as a measure of the "satisfaction" of a node in his group of friends and the contribution is a measure of how much an individual contributes to his group of friends with a resource (if you have a group of friends you almost always contribute with a resource, the most basic is time). Moreover, strategies change only when a node is not satisfied enough or when he has not enough friends, similarly to what happens to group of friends in real-life.

Differently to other gossiping strategies known in literature, like the "Copy-and-Rewire" algorithm, we don't assume that an unsatisfied node could randomly meet one that is really satisfied, instead we merge the view of the most unsatisfied, because the most unsatisfied should be those that commit in order to find new friends, and should only use their knowledge.

4 Experimental framework, Parameter selection and results

The experimental framework is NetLogo, [3] an environment for agent-based modeling. The experiments were carried out by considering the three network dynamics algorithm for each model baseline, for which the number of nodes were chosen to be $10, 50, 100$ but further experiments can be performed by tweaking the parameters in the provided NetLogo file. In addition to the already present stopping criteria, the experiments terminate after a fixed amout of "ticks" where a tick can be tought as a single iterate of each algorithm, in this case the value is 2000, which corresponds to about one minute if the simulations are performed using normal speed. Apart from the number of friends, the number of friends of friends and the mean average number of friends of friends, it is taken into account:

- The clustering coefficient, which is a measure of the triadic closure of the nodes in the graph, it is needed because it is a good estimate of how the friends of a node are also friends of each other.
- After the termination of each algorithm, the fraction of nodes in the giant component with respect to the removed nodes is taken into account, this is useful for two reasons: it gives a visual estimate of how robust the outcomes of these algorithms are and it is useful to understand if there were more components in the network if the view becomes occluded. This measure is considered both for random uniform node removal and preferential removal (the nodes with the highest degree are removed first).
- In the same setting of the measure above, it is taken into account the average path length with respect to the number of nodes removed, but only for networks with number of nodes equal to 50, since the complexity of the Dijkstra's Algorithm is $O(|E| + |V|log|V|)$ and we have to call it for each pair of nodes.

Since the goal is to model how friendships are made between individuals, we do not require a high `edge-probability`, by leaving it between 0 and 0.2 it is possible to obtain graphs were few relationships are build, so that the variance is not too high. Experiments were carried out for `edge-probability` $= 0.05$. By incrementing `edge-probability` it is expected a faster convergence of the algorithms, given the stopping criteria.
For the random dynamics algorithm `rewire-probability` is kept in the same range of `edge-probability` because by incrementing it above 0.2 it would be observable a higher (vertical) distance between the curve related to the mean number of friends and the curve related to the mean average number of friends of friends. The experiments were carried out using `rewire-probability` $= 0.2$}. For both models for network configurations, by incrementing the number of nodes we observe a higher distance between the average number of friends and the average mean number of friends of friends.

---

[3] NetLogo programming language: https://en.wikipedia.org/wiki/NetLogo

The plots about the average number of friends and average mean number of friends of friends result in highly oscillating values, since this algorithm is purely random. The clustering coefficient is generally low, as expected since we are rewiring all the nodes randomly (and with a low probability), so the probability of having a triadic closure is expected to be low consequently. While performing tests about robustness, the average path length and fraction of nodes in the giant component decay steadily because we obtain a single component. The robustness plots of the two baselines are fairly similar, since the algorithm rewires the previously existing nodes, however with the Erdős-Rényi baseline we obtain a network which is slightly more robust.
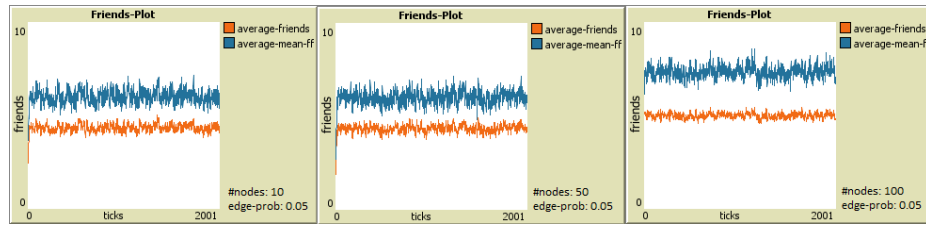
Fig. 1. Statistics about the simple random dynamics algorithm with the Erdős-Rényi baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
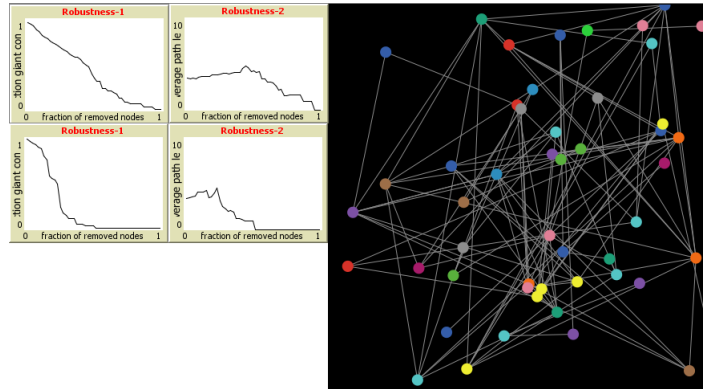
Fig. 2. plots about the robustness of the graphs obtained after running the simple random dynamics algorithm with the Erdős-Rényi baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes).
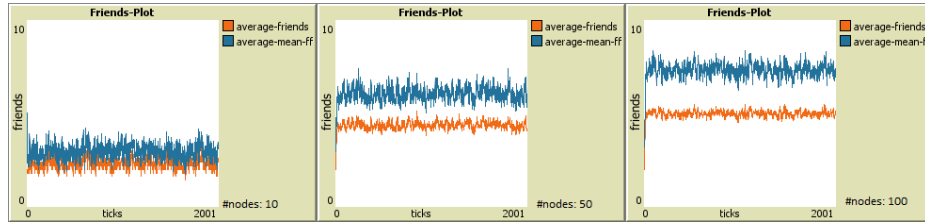
Fig. 3. Statistics about the simple random dynamics algorithm with the Barabási–Albert baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
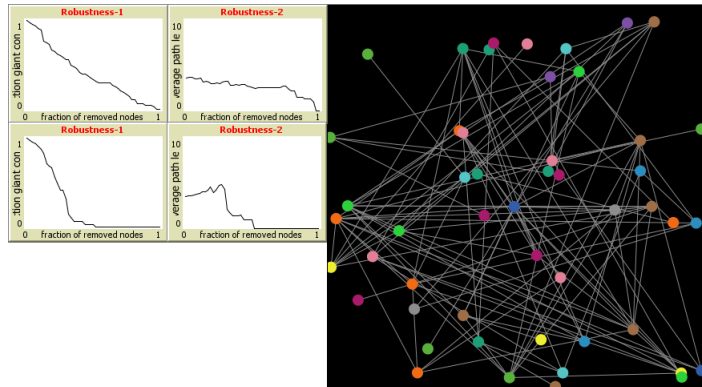

Fig. 4. plots about the robustness of the graphs obtained after running the simple random dynamics algorithm with the Barabási–Albert baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes).

The genetic algorithm was created in order to achieve low variance. Even if there is some noise and the outcome is a single giant component (so robustness is kept, but a giant component is not a good model for friendships, having different components is more appropriate), the plots show how the distance between the average number of friends and the average mean number of friends of friends is overall very low. The curves related to the average number of friends and the average number of friends of friends tend to stay in nearby the value $n/2$ (where n is the number of nodes), while with the previous algorithm only the curve related to the average number of friends exhibited that behaviour. The clustering coefficient is almost always below 0.1, meaning that the friends of the nodes in this kind of network do not tend to connect to each other, as expected since the fitness function prioritizes nodes with a lower degree. By observing the robustness plots it is possible to observe that with both baselines we obtain networks that are more robust with respect to the simple random dynamics, furthermore, even in this case, using the Barabási–Albert baseline result in networks that are (slightly) more robust, since the average path length and the fraction of nodes in the giant component decay more smoothly.
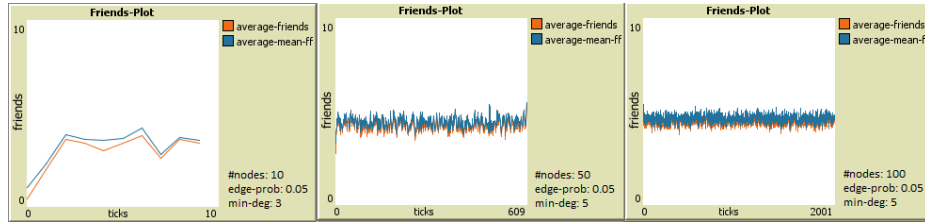

Fig. 5. Statistics about the genetic algorithm with the Erdős-Rényi baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
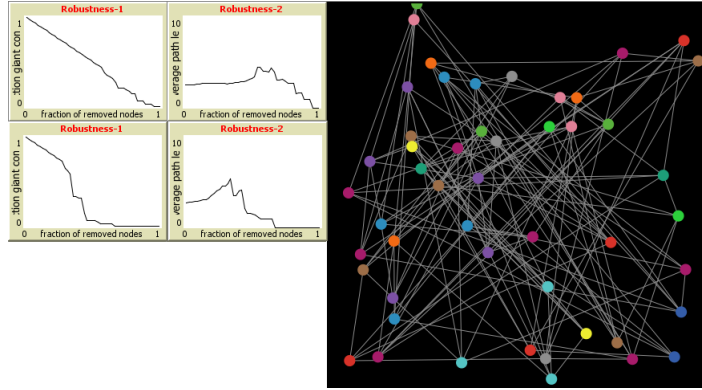
Fig. 6. plots about the robustness of the graphs obtained after running the genetic algorithm with the Erdős-Rényi baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes)
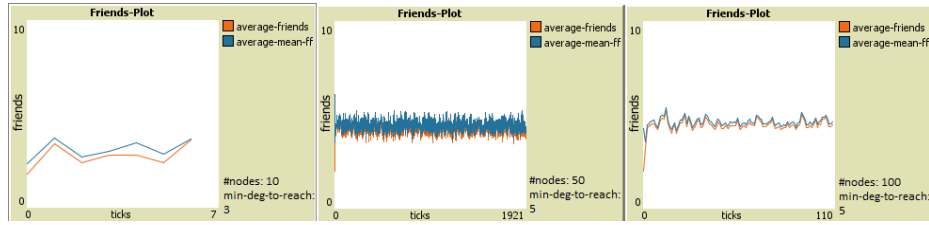

Fig. 7. Statistics about the genetic algorithm with the Barabási–Albert baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
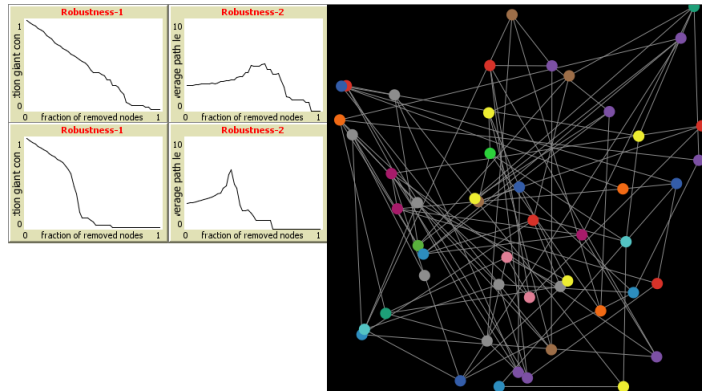

Fig. 8. plots about the robustness of the graphs obtained after running the genetic algorithm with the Barabási–Albert baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes)

For the game-theory based algorithm the choice of the parameters is done as following:

- the number of gossiping nodes is 2, it is the value for which small clusters are formed, empirically it works well with both models and with different amount of nodes, incrementing this value brings almost always to the formation of a single giant component.
- `y-max` is chosen to be 12, it is an arbitrary value, it was chosen because it represents the number of hours in half a day.
- `b-abstain` and `b-collaborate` are chosen to be respectively 0.6 and 1.2, this because we want to penalize 40% more those nodes who do not sign the treaty and we want to give 20% more to those who sign. Again, this is arbitrary but works better than other proportions in terms of how the nodes are chosen.
- `gain-percentage` is chosen to be 110%, nodes are rational, but they should not need huge improvements in order to make friends.
- `drop-prob` is chosen to be 0.6 for the same reason of the probability of dropping a node in the genetic algorithm, however 0.4 led to a higher variance and the formation of a giant component most of the times, while with 0.6 it is possible to observe some nodes that cluster themselves apart or stay lonely until the timeout is reached.
- `payoff-threshold` is chosen to be 4, less than half the value of `y-max`, again this value is arbitrary but reasonable in the context of the number of hours that a node could give as contribution.
- `num-neigh-threshold` is chosen to be 2 because lonely nodes must change strategy in order to try to achieve better payoffs (and so friends).

The main characteristic of this algorithm is that it allows the formation of small clusters with both the baselines, thus the clustering coefficient is higher with respect to the previous algorithms (> 0.5), moreover the average number of friends and the mean average number of friends of friends are the lowest with respect to all the algorithms and for the Barabási–Albert baseline we have a steep decay of the mean average number of friends of friends during the initial iterates (because we consider the nodes with the lowest degree first and it is very likely to destroy edges with those nodes that have the highest degree, because of the preferential attachment). Again, the starting baseline does not influence so much the outcome of the algorithm, which consists in graphs that are less robust with respect to those obtained with the previous algorithms (as expected, since by having small clusters the fraction of nodes in the giant component and the average path length are low from the start).
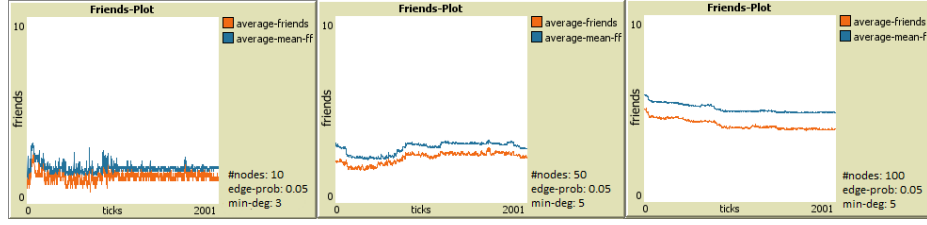
Fig. 9. Statistics about the game theory based algorithm with the Erdős-Rényi baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
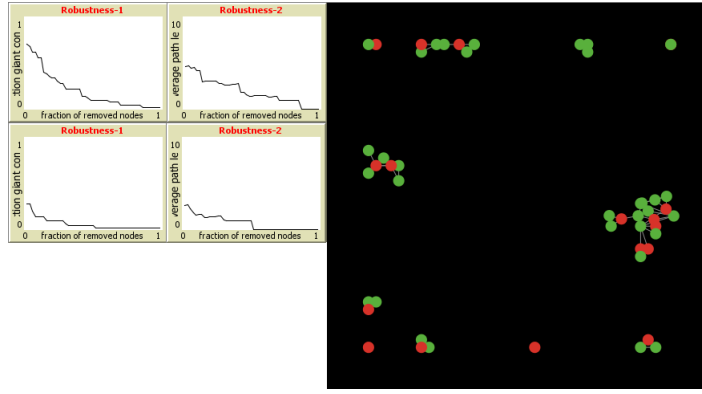

Fig. 10. plots about the robustness of the graphs obtained after running the game theory based algorithm with the Erdős-Rényi baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes)
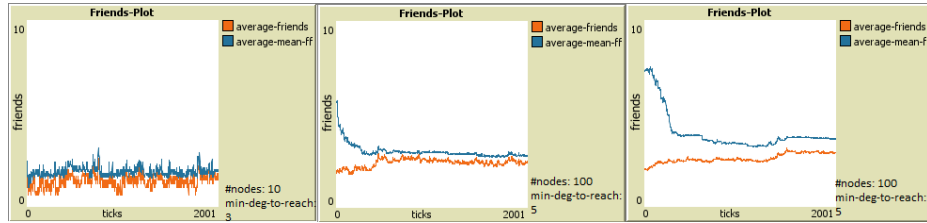

Fig. 11. Statistics about the game theory based algorithm with the Barabási–Albert baseline. The three plots show the average number of friends and the average number of friends of friends for increasing number of nodes.
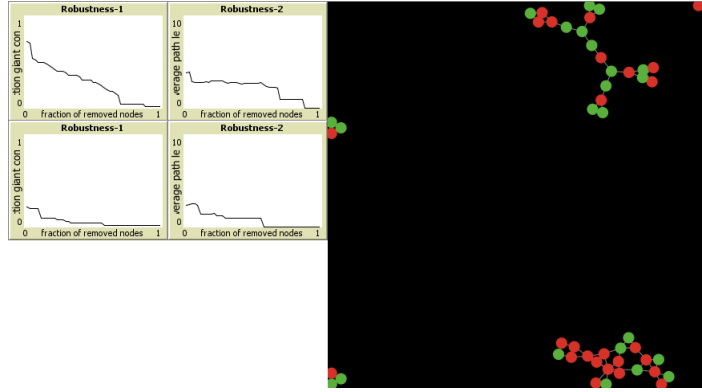
Fig. 12. plots about the robustness of the graphs obtained after running the game theory based algorithm with the Barabási–Albert baseline. The two plots on top refer to uniform random removal, while the plots on the bottom refer to preferential removal. On the right there is an example of the obtained graphs (with 50 nodes)

5   Conclusions

The objective of this paper was to design algorithms for network dynamics and analyze their outcome after starting from different baselines. The simple random dynamics algorithm is a basis of comparison for the two other algorithms, as its behaviour is purely chaotic and no particular final configuration is expected to be observed. The genetic algorithm is designed with the objective of obtaining low variance by prioritize the formation of edges with those nodes that have the lowest degree, while considering the characteristics of each node which were encoded as chromosomes (that can also mutate or change completely). The Treaty Game based algorithm is the most sophisticated one as it allows the formation of small clusters and allows all the agents to perform rational decisions about what edges to create and what to destroy, at the price of a lower robustness. No peculiar differences emerged from using a baseline with respect to the other one, as the algorithms here presented (except for the random one) consider from the start those nodes that have the lowest degree.

References

1. Scott L Feld. Why your friends have more friends than you do.
   American journal of sociology, 96(6):1464-1477, 1991.
2. Feld, Scott L., and Bernard Grofman. Variations in Class Size, the Class Size
   Paradox, and Consequences for Students. *Research in Higher Education* 6:215-22,
   1977.
3. Scott Barrett. Coordination vs. voluntarism and enforcement in sustaining international environmental cooperation.
   Proceedings of the National Academy of Sciences, 113(51):14515-14522, 2016.