CS 7280, Fall 2017: Project Report

The RP-model: Properties and Variations

Ankit Srivastava

March 15, 2018

Abstract

Preference for reusing existing modules over building the new modules from scratch can explain the "hourglass effect" in hierarchical systems. *Reuse Preference* model (RP-model) has been proposed for generating hourglass networks. In this project, I attempted to explore if the amount of preference for reuse required for the networks generated by the model to exhibit the hourglass property depends on the size of the network. The preference for reuse in hierarchical systems can likely be attributed to the costs associated with every new connection in such systems. Systems trying to minimize the cost associated with the new modules will end up reusing more existing modules. For the second part of the project, I explored a modification of the RP-model in which the number of connections in the network depends on the preference for reuse and showed that this modification generates hourglass-like networks even with a lower preference for reuse.

1 Introduction

As early as 1962, Nobel laureate Herbert Simon hypothesized that various complex systems encountered in nature are constructed from smaller simpler subsystems in a hierarchical manner [1]. Many years later, in 2003, Ravasz & Barabási developed a hierarchical network model and used it to show that multiple real-world networks are indeed hierarchically modular [2]. Since then, further studies of real-world hierarchically modular systems have shown that such systems tend to have a small number of highly conserved intermediate modules on the path of conversion from numerous elementary sources to diverse complex targets [3, 4, 5, 6, 7]. The dependency networks of these systems resembles an hourglass, with the highly conserved modules forming the "waist" of the hourglass. Therefore, these dependency networks are commonly referred to as *hourglass networks*.

Sabrin & Dovrolis hypothesized that hourglass dependency networks are observed in hierarchical systems in which new modules have a preference for reusing the modules closer to them in the hierarchy over the modules that are closer to the source modules [8]. They proposed the *Reuse Preference* model (RP-model) for generating dependency networks which exhibit the hourglass property, parameterized by a *reuse parameter* α. They observed that for positive values of α, the dependency networks generated by the model show pronounced hourglass structure. For quantifying how close to the hourglass structure any given dependency network is, they defined the *Hourglass score* (*H-score*) ($0 \leq H < 1$) which is close to 1 for an ideal hourglass-like network (i.e., a network with just one vertex in the "waist").

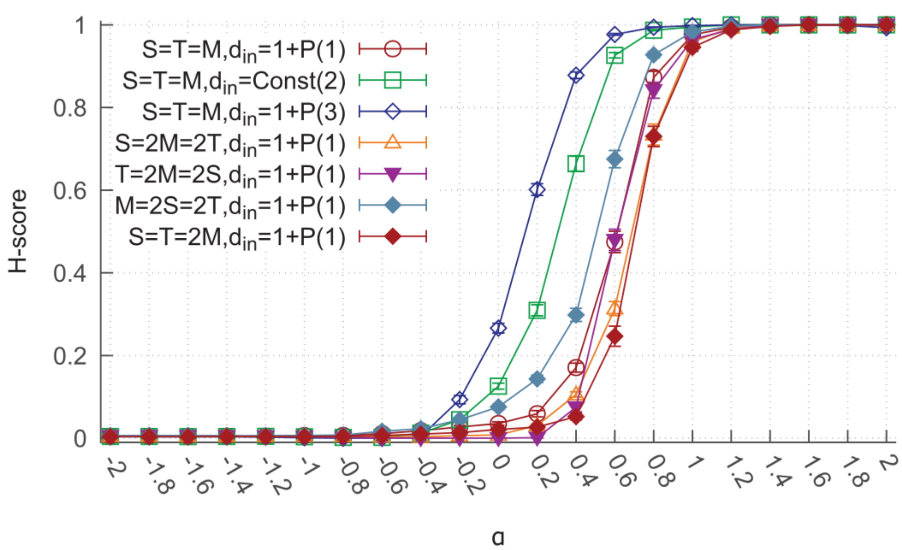

Figure 1: H-score as a function of α. [8]

The plots of H-scores for the networks generated using the RP-model, with different configurations, as a function of α is shown in Figure 1. From the figure, it can be seen that the H-scores are close to 0 when α is negative for all the different configurations. As α increases, there seems to be a *phase-transition* as the H-scores jump to a value close to 1 for very small changes in the value of α. If there is indeed a phase-transition, then it would mean that small changes in the characteristics of fragile systems, those in which source modules and target modules are tightly coupled, can lead to more evolvable systems in which source and target modules are decoupled. Therefore, for the first part of this project, I attempted to confirm the existence of this apparent phase-transition.

The formulation of the RP-model in [8] assumes that the number of existing modules that a newly added module uses does not depend on the modules that it ends up using. However, intuitively, if the new module uses modules which are lower in the hierarchy, then it would need to use more such elemental modules since it would need to build more "functionality" from scratch. Therefore, for the second part of this project, I developed an alternate formulation of the RP-model in which the number of modules used by a new module is dependent on the number of source modules and the reuse parameter and then compared the hourglass properties of networks generated using the alternate formulation of the model with the corresponding properties for the original model.

In this report, I will first provide some background on the formulation of the original RP-model from [8] and the revised formulation of the model used in this project. This is required for discussing the experiments using the model and the modification of the model in this project. Then, I will briefly discuss some related work. Finally, I will describe the methodology that I followed for this project followed by a discussion of the results and scope for future work.

2 Background

The RP-model proposed in [8] generates directed acyclic networks which represent dependency networks for hierarchical systems. The model is parameterized using the following parameters:

- V - Total number of vertices in the network.
- S, M, and T - Number of source, intermediate, and target vertices in the network. These represent elementary source modules, modules of varying complexity at intermediate levels of the hierarchy, and the final products of the system. ($V = S + M + T$)
- $\mathrm{d_{in}}$ - The in-degree of every vertex in the network. This is supposed to represent the number of dependencies of modules in the network. In [8], this was either constant or picked from a Poisson distribution.
- α - Reuse parameter.

The generation of any network using the model is done in $M+2$ iterations. In the first iteration, all the source vertices are created in a batch. The intermediate vertices are created one at a time in M subsequent iterations. Finally, all the target vertices are created in the last iteration. In the first step of every iteration, any existing vertices are assigned ranks such that the vertex created in the previous iteration is assigned rank 1, the one created in the iteration before that is assigned rank 2, and so on. If multiple vertices were created in the same iteration, then those vertices are assigned ranks randomly in the corresponding range. Using the assigned ranks, the probability of a connection

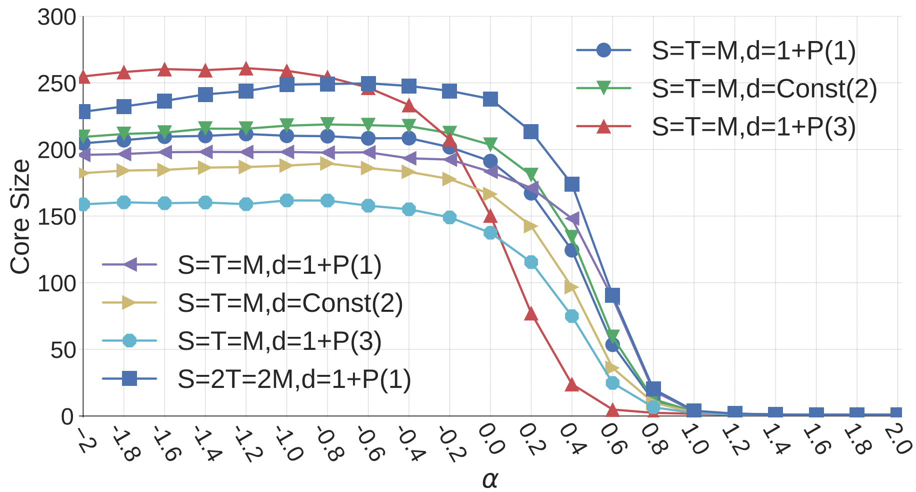

Figure 2: Core size as a function of α with the revised model formulation.

from an existing vertex u with rank r_u to the vertex v created in the current iteration is calculated as

$$\text{Prob}[(u, v)] = \frac{r_u^{-\alpha}}{\sum r_i^{-\alpha}}$$

$d_{\text{in}}(v)$ connections are then created for every v by picking the origin vertices using the probabilities calculated above. It is trivial to observe that for negative α, older vertices are more likely to get picked while for positive α, the connections will more likely be created from newer vertices.

The model described above creates an unweighted dependency network. However, the results discussed in this project were obtained using a revised formulation of the RP-model provided by the authors of [8] which generates a weighted dependency network. In the revised formulation, the first intermediate vertex is created differently from all the other vertices. Instead of creating connections from d_{in} sources, connections are created from all the sources to the vertex, each with weight d_{in}/S. All the other vertices are created as before, and the weight of all the other connections is set to 1. I implemented this revised formulation and repeated the experiments in [8]. The plots for the effect of α on core sizes and H-scores with the revised model are shown in Figure 2 and Figure 3. Hereafter, in this report, this revised formulation of the model is referred to as the RP-model.

3 Related Work

Multiple network models in the literature have been shown to exhibit phase-transition i.e., small changes in one or more model parameters generate net-

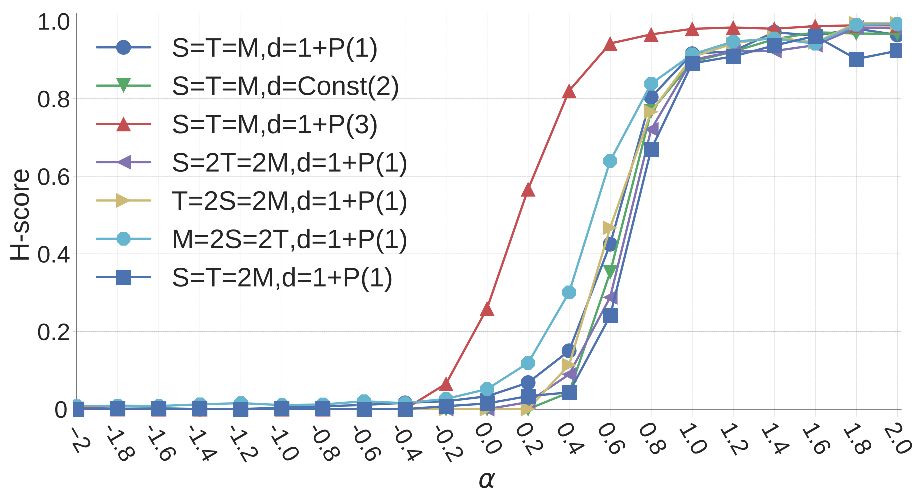

Figure 3: H-score as a function of α with the revised model formulation.

works with very different properties. Erdős & Rényi showed that for a very small increase in the average vertex degree, the vertices in a random graph go from being disconnected to forming a giant connected component [9]. Networks with a power-law degree distribution exhibit scale-free properties when the degree exponent in the power-law is less than 3 [10]. Optimization based models [11] generate star, scale-free, or exponential networks for different values of the weighting parameter [12].

A few studies have shown that the growth of hierarchy and modular structure in networks is due to a cost attached with every connection [13, 14]. This is because the cost of building the network can be reduced by connecting to a relatively smaller number of complex modules, instead of connecting to numerous, relatively general, sources. The original formulation of the RP-model, however, doesn't consider the costs of connections. Considering this factor could better explain the emergence of hourglass effect in the dependency networks of real-world systems.

4 Methodology

There were two main components of the work proposed as part of this project. The first was to investigate the existence of the apparent phase-transition in H-scores of the networks from 0 to a value close to 1 as α used for generating the networks is increased. I explored two approaches for this purpose. The first was to use the mathematical formulation of the RP-model to theoretically show the existence, or lack thereof, of the transition. However, I didn't make any significant headway using this approach for a while. Therefore, given the time bounded nature of this project, I proceeded with the alternate approach of using simulations. For this purpose, I varied the parameters of the model and noted the changes in the properties of the generated networks. The results obtained from the simulations are presented and discussed in Section 5.

For the second part of the project, I investigated the effect that making the in-degree distribution dependent on the preference for reuse has on the hourglass network properties. Intuitively, when the preference for reuse is low, any newly added module is likely to depend on almost all the source modules. On the other hand, when the preference for reuse is high, the newly added module is more likely to be dependent on a smaller number of intermediate modules. With this intuition, I proposed a new function for choosing the in-degrees of newly added module. The function is dependent on the number of sources in the system (S) and the minimum number of connections any new module should have (k) and is shown below.

$$f(S,k) = S\left(\frac{k+e^{-\alpha}}{S+e^{-\alpha}}\right)$$

This function has the desired properties i.e., $f(S,k) \to S$ as $\alpha \to -\infty$ and $f(S,k) \to k$ as $\alpha \to \infty$. The value of the function for different values of S and k was plotted for $-10 \le \alpha \le 10$ on a semi-log plot and is shown in Figure 4. It can be seen that the plots of $f(S,k)$ in the negative α region for the same S values overlap for very small α but start to separate when $\alpha \approx -4$ and the transition seems to be complete around $\alpha \approx 1$. Then, for positive α, the plots for the same values of k overlap. The transition in the negative α region close to $\alpha = 0$ suggests that the phase-transition observed in the H-scores should become more pronounced with this in-degree function. The results of the experiments performed using this function are discussed in Section 5.

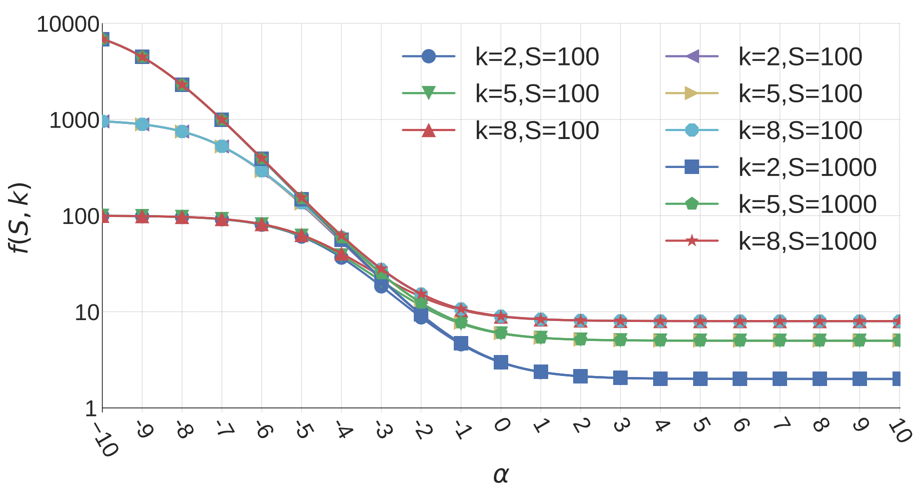

Figure 4: Plots of $f(S,k)$ as a function of α for different values of S and k.

5 Results & Discussion

I implemented the RP-model described in Section 2 and validated it using the implementation provided by the authors of [8]. The implementation can be accessed on Github [15] and was used for all the experiments reported here.

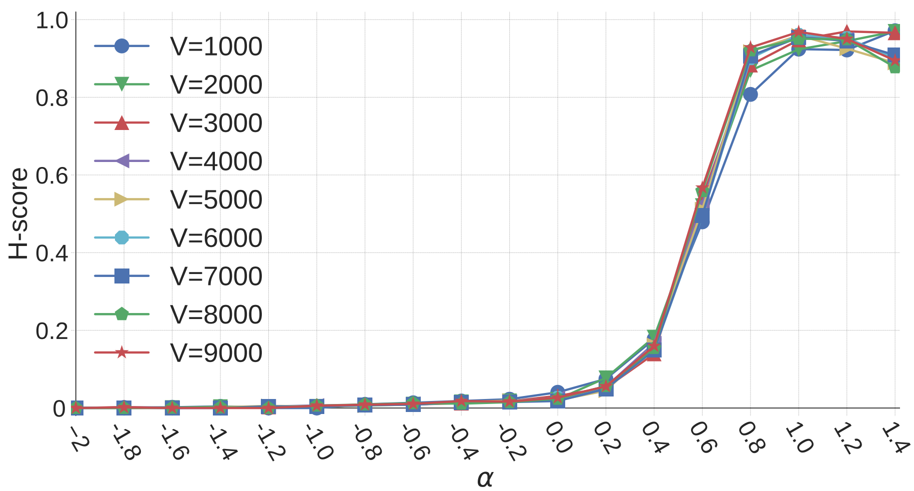

Figure 5: H-score as a function of α for different values of V and "$S = M = T$,$\mathrm{d_{in}}$=$1 + P(1)$" configuration.

In order to investigate the existence of phase-transition, I chose the configuration "$S = M = T$,$\mathrm{d_{in}}$=$1 + P(1)$" from [8]. In this configuration, the number of source, intermediate, and target vertices in the network are equal and same as one third of the total number of vertices in the network. Further, the number of dependencies of any new vertex is set using the Poisson distribution with the parameter value of 1. For this configuration, V was varied between 1000 and 9000, in steps of 1000, and α was varied between -2.0 and 1.4, in steps of 0.2. Every experiment was repeated 10 times, with different pseudo-random number generator seeds, and the reported values were averaged. The resulting plot is shown in Figure 5.

It can be seen that the plots for different values of V follow the same curve and there are no significant changes in the shape of the curve as the value of V is increased. This is further confirmed by Figure 6 which shows the plots in Figure 5 only for the transition region. Few simulation runs for other configurations, not shown in this report, seem to follow the same trend, i.e, increasing V does not seem to change the shape of the curve. This seems to indicate that there is no phase-transition, since if there was phase-transition then the transition from low to high H-scores would've been more pronounced with increasing V. However, more experiments are required for making a conclusive statement.

For gaging the effect that $f(S, k)$ has on the hourglass properties, I repeated the experiments in [8] with the following change. The in-degree was set using

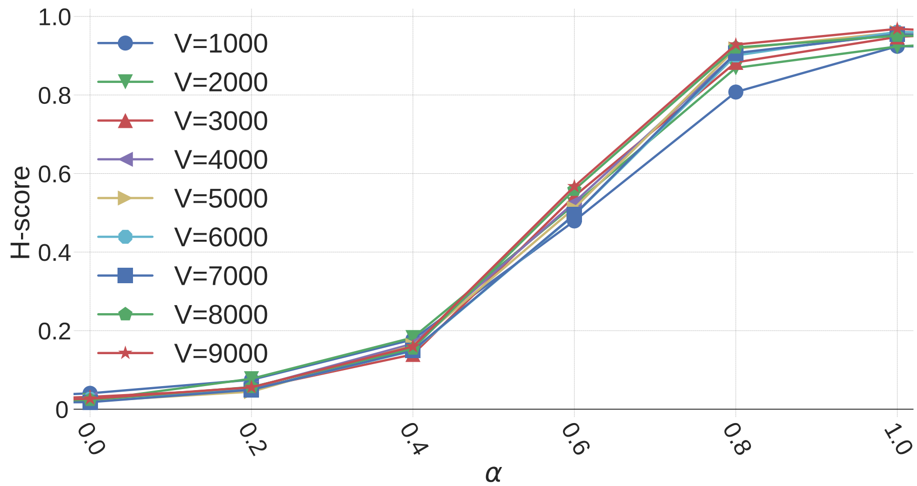

Figure 6: Plots in Figure 5 shown just for $\alpha \in [0.0, 1.0]$.

$f(S, \mathrm{d_{in}})$, where $\mathrm{d_{in}}$ was the degree in the original configuration. Every experiment was repeated 20 times for $V = 1000$, with different pseudo-random number generator seeds, and the reported values are an average of the results thus obtained. The plots of the H-scores as a function of α are shown in Figure 7.

The dip in the H-scores when $\alpha = 0.8$ is due to the definition $f(S, k)$. Since the in-degrees are integral, the in-degrees are set to $\mathrm{round}(f(S, k))$, instead of the floating-point value of $f(S, k)$. The plot of $\mathrm{round}(f(S, k))$ as a function of

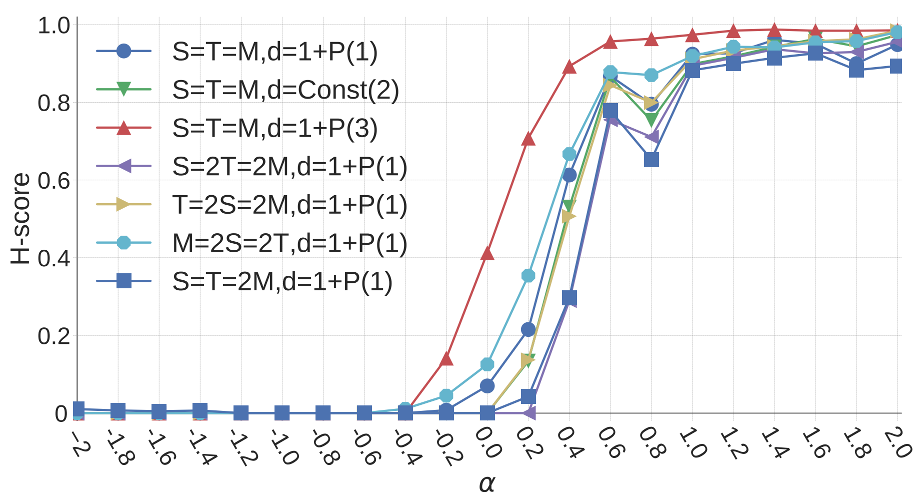

Figure 7: H-score as a function of α with the in-degrees set to $f(S, \mathrm{d_{in}})$.

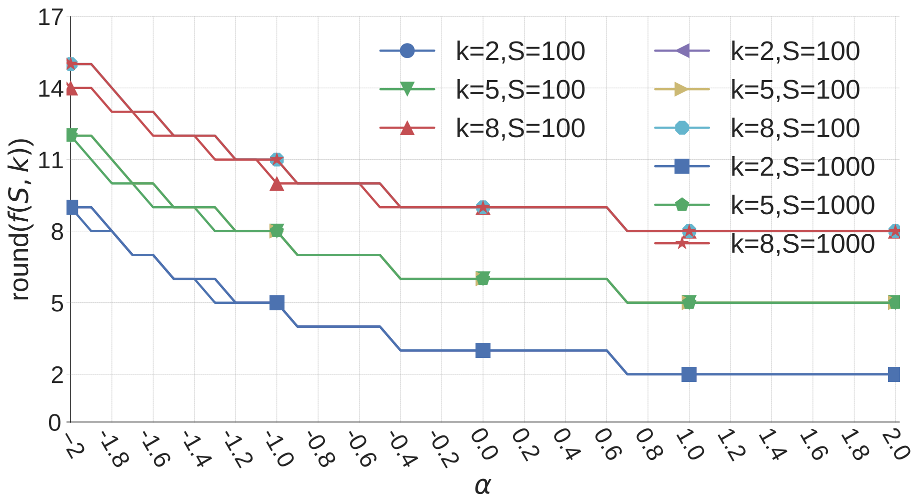

Figure 8: Plots of $\text{round}(f(S,k))$ as a function of α.

α is shown in Figure 8. It can be seen that the value of $\text{round}(f(S,k))$ drops by 1 for all S and k combinations in the α interval $(0.6, 0.8)$. Therefore, the dip in H-scores can be attributed to the decrease in the in-degree of the vertices in the generated networks.

Figure 9 shows the comparison between the H-scores for the networks generated with the in-degree function and without it, for only three configurations for the purpose of clarity. The plots of the H-scores for the networks generated

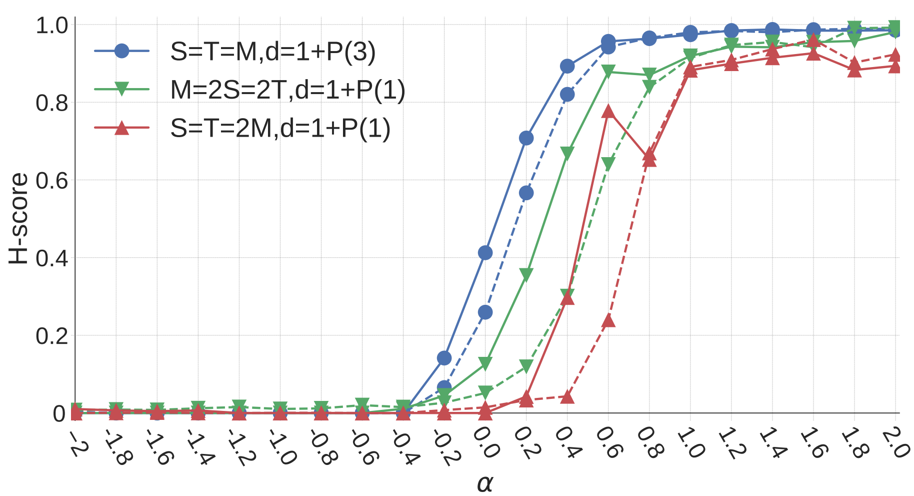

Figure 9: Comparison of the H-scores in Figure 3 and Figure 7.

without the in-degree function are shown with dashed lines while those for the networks generated using the in-degree function is shown with solid lines. As hypothesized in Section 4, the transition from low to high H-scores is steeper with the in-degree function. This confirms the initial hypothesis that attaching a cost to every connection in the RP-model can better explain the emergence of hourglass properties in the generated networks.

6 Future Work

The scope for future work in this project is quite extensive. The implementation of the analysis of networks for determining the hourglass properties is prohibitively slow for networks with a large number of vertices ($V > 9000$) and takes a long time even for networks with smaller number of vertices. This was the main hindrance to running multiple experiments for exploring the phase-transition aspect of this project. I believe that a more optimized implementation would aid this effort. Since the current implementation is written in Python, this may be as trivial as using a language like C or C++ for the purpose. Further, using the mathematical formulation of the model for the purpose may void the need for simulations altogether.

There is also room for improvement in the formulation of the in-degree function. As discussed in Section 5, the rounded value of the function dips in the α interval $(0.6, 0.8)$. It may be pertinent to investigate this further and modify the function such that this dip happens in some interval where $\alpha < 0$ so that the effect of the dip on the H-scores can be avoided.

References

[1] H. A. Simon, "The architecture of complexity," *Proceedings of the American Philosophical Society*, vol. 106, no. 6, pp. 467–482, 1962.

[2] E. Ravasz and A.-L. Barabási, "Hierarchical organization in complex networks," *Physical Review E*, vol. 67, no. 2, p. 026112, 2003.

[3] R. Q. Quiroga, L. Reddy, G. Kreiman, C. Koch, and I. Fried, "Invariant visual representation by single neurons in the human brain," *Nature*, vol. 435, no. 7045, pp. 1102–1107, 2005.

[4] J. Supper, L. Spangenberg, H. Planatscher, A. Dräger, A. Schröder, and A. Zell, "Bowtiebuilder: modeling signal transduction pathways," *BMC systems biology*, vol. 3, no. 1, p. 67, 2009.

[5] A. T. Kalinka, K. M. Varga, D. T. Gerrard, S. Preibisch, D. L. Corcoran, J. Jarrells, U. Ohler, C. M. Bergman, and P. Tomancak, "Gene expression divergence recapitulates the developmental hourglass model," *Nature*, vol. 468, no. 7325, p. 811, 2010.

[6] M. Quint, H.-G. Drost, A. Gabel, K. K. Ullrich, M. Bönn, and I. Grosse, "A transcriptomic hourglass in plant embryogenesis," *Nature*, vol. 490, no. 7418, p. 98, 2012.

[7] S. Akhshabi and C. Dovrolis, "The evolution of layered protocol stacks leads to an hourglass-shaped architecture," in *Dynamics On and Of Complex Networks, Volume 2*, pp. 55–88, Springer, 2013.

[8] K. M. Sabrin and C. Dovrolis, "The hourglass effect in hierarchical dependency networks," *Network Science*, pp. 1–39, 2017.

[9] P. Erdos and A. Rényi, "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, vol. 5, no. 1, pp. 17–60, 1960.

[10] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Reviews of modern physics*, vol. 74, no. 1, p. 47, 2002.

[11] A. Fabrikant, E. Koutsoupias, and C. Papadimitriou, "Heuristically optimized trade-offs: A new paradigm for power laws in the internet," *Automata, languages and programming*, pp. 781–781, 2002.

[12] A.-L. Barabási, *Network science.* Cambridge university press, 2016.

[13] J. Clune, J.-B. Mouret, and H. Lipson, "The evolutionary origins of modularity," in *Proc. R. Soc. B*, vol. 280, p. 20122863, The Royal Society, 2013.

[14] H. Mengistu, J. Huizinga, J.-B. Mouret, and J. Clune, "The evolutionary origins of hierarchy," *PLoS computational biology*, vol. 12, no. 6, p. e1004829, 2016.

[15] A. Srivastava, "RP-model." `https://github.com/asrivast28/rp-model`, 2017.