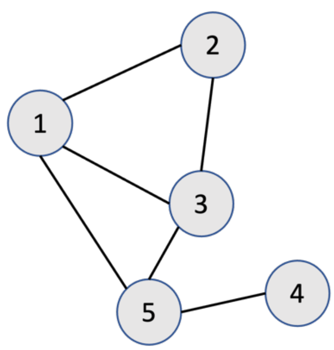
Marko Suchy
4/19/24
PHYS 345
Final Paper

Entropy in Complex Systems

‘Complex systems’ by themselves are not particularly easy to define. The field of complex systems is rather new, becoming prominent around the turn of the century. The field is largely interdisciplinary, and contributions come from literature across sociology, ecology, condensed-matter physics, computer science, economics, and beyond. While there is no technical definition of a complex system, Mark Newman describes them as “a system composed of many interacting parts, such that the collective behavior of those parts together is more than the sum of their individual behaviors.” [1]

In the study of complex systems, it is useful to represent our parts, or ‘agents,’ and their interconnections as networks. This gives us a mathematical framework from which to work: graph theory. In graph theory, a graph G, can be represented by its adjacency matrix A, where A is an $N \times N$ matrix, with N as the number of nodes (agents) in the system. A 1 at matrix element a_{ij} indicates a relationship between nodes i and j, as seen in Figure 1. The power of graphs is their ability to represent any sort of relational data. Thus, they can be used to wide range of systems across disciplines.



$$A_{i,j} = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{bmatrix}$$

Figure 1: A simple example of an unweighted, undirected graph G, with adjacency matrix A.

The literature on network analysis widespread.[1] Several techniques and algorithms exist in graph theory which may offer insights into the nature of complex systems. There are many standard metrics which quantify the connectedness, centrality, and other characteristics of networks, at both the individual node and entire system levels. When applying entropy to networks representing complex systems, we measure the uncertainty in these metrics. While the formulation of entropy varies with respect to the particular metric which is being examined, all entropy formulations are derivate of Shannon’s Entropy, as seen in Equation 1.

[1] One issue with network science is the siloed nature of literature. While network analysts across disciplines often use similar methods (centrality, percolation, community detection) they rarely work together to generalize their findings to complex systems at large.

$$H(X) = -\sum_{i=1}^{n} p_i \, log_2 \, p_i$$

Equation 1: A general formulation of Shannon's Entropy as presented in [3] with units of bits, where H(X) is the Shannon's entropy of ensemble X, and p_i *is the probability of a particular microstate occurring.*

Shannon's entropy as presented in [3] was originally proposed as the lower limit of data compression and is foundational in information theory. It is very similar to the Gibbs expression of entropy, just lacking the unit-giving constant out front. In the formulation of Shannon's entropy shown in Equation 2, the base-2 of the logarithm function gives the informational unit of bits. Because the probability mass function (p_i) is not well-defined physically, the Shannon's entropy is not a well-defined physical state function like the Gibbs entropy. Thus, derivatives of the Shannon's entropy are not physically interpretable like the Gibbs entropy, but more generally a measure of the sensitivity entropy is to changes in particular probabilities. When applying the Shannon's entropy to complex systems, we take distributions of node and system level metrics as the probability mass function.

Traditionally, graph information entropy has been based on graph level metrics, like number of nodes, edges, or degree distribution, which can be used to measure the structural entropy of a graph. Other types of information entropy have been developed which focus on the entropy of a particular node in a network, examining distributions of neighbor degree, strength (an extension of degree for weighted networks,) and several types of centralities. These different formulations of entropy in complex systems have been applied widely throughout literature, on a plethora of different projects. In their survey of information entropy metrics for complex systems, Omar and Plapper find that applications of entropy in complex networks are scattered throughout over 10 journals, and many disciplines [2].

In 2017, Wang et al. considered technology project portfolios of large enterprises as a complex network of interacting agents who, much like biological agents, sometimes work together and sometimes compete for the same resources [4]. Wang et al. represent projects as nodes in networks, and assign weighted edges[2] based on software systems shared between projects. The researchers use the Louvain community detection algorithm to group nodes by sharing resources, information, or objectives. They calculate the efficiency and risk of communities using entropic methods, then search for the optimal community structure to minimize risk at a certain efficiency threshold.

For assessing efficiency, the Wang et al. use a formulation of Shannon's entropy based on the size of the community, which is meant to capture the integration efficiency of new technology, and a sum of clustering coefficient and closeness centrality, which is a measure of the cooperation levels within the community. The underlying assumption for this efficiency entropy is that information spreads well through small, well-connected communities.[3] For risk entropy, Wang et al. consider the weighted degree of individual nodes, as a measure of homogeneity in the software products used in the community. The underlying assumption here is that more homogeneity induces more risk. The formulation for risk entropy used by Wang et al. can be seen in equation 2. Efficiency entropy is quite a bit more complicated thus it is excluded

[2] See Equation 1 from [4] for a formal definition of edge weights.
[3] This is closely related to the conclusion of my honors thesis! It's cool to see how all these things connect.

from this paper for brevity, but it can be seen in equations 4-7 of [4]. Ultimately Wang et al. suggest that their model could be well utilized by project portfolio managers to quantify and minimize risk of their IT project portfolios.

$$\mathrm{H_R} = \sum_{\mathrm{j}=1}^{\mathrm{s}} \frac{\mathrm{n_j}}{\mathrm{n}} \mathrm{H_{RM_j}} \;\; ; \;\; \mathrm{H_{RM_j}} = \sum_{\mathrm{i}=1}^{\mathrm{j}} \mathrm{p_i} \ln \mathrm{p_i} \;\; ; \;\; \mathrm{p_i} = \frac{k_i^w}{\sum_{\mathrm{i}=1}^{\mathrm{n_j}} k_i^w}$$

Equation 2: Wang et al.'s formulation for risk entropy, $\mathrm{H_R}$*, where projects j through s are in community j,* n *is the number of projects in the entire system,* $\mathrm{n_j}$ *is the number of projects in community j, and* $\mathrm{k_i^w}$ *is the weighted degree of node i.*

Clearly, entropy in complex systems can be useful in understanding micro and macro level system dynamics. For a comprehensive review of how Shannon's entropy has been used to understand complex networks, see [2]. While this entropy is not physically interpretable as entropy from thermodynamics, it draws strong parallels with the Gibbs formulation. As Omar & Plapper say, the field is young, and there are still many interesting network-based formulations of Shannon's entropy to explore.

Bibliography

[1] Newman, M. E. (2011). Complex systems: A survey. *arXiv preprint arXiv:1112.1440*.

[2] Omar, Y. M., & Plapper, P. (2020). A survey of information entropy metrics for complex networks. *Entropy*, *22*(12), 1417.

[3] Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379-423.

[4] Wang, Q., Zeng, G., & Tu, X. (2017). Information technology project portfolio implementation process optimization based on complex network theory and entropy. *Entropy*, *19*(6), 287.