Examining the Variability in Network Populations and its role in Generative Models: Supplementary Materials

For all the experiments conducted in Section 4, the single observed network G^* is randomly chosen among the networks in dataset. However, the pairwise comparison relies on the chosen G^*, and arbitrary choice of G^*, e.g. an outlier, could cause misleading results. In this section, the impact of choice of G^* on the results presented in Section 4 are analyzed. In summary, for the 9 datasets except for autonomous systems, the three tested network metrics are consistent regardless of choice of G^*. For autonomous systems, the difference of variability between network populations generated by true process and generative models is more significant than the difference caused by choice of G^*.

Firstly, we compute the dissimilarity of degree distribution, local assortativity and local transitivity with respect to all networks $\{G \in \mathcal{G}_{A^*}\}$. Specifically, for each network G_i we compute the KS statistics D_i of the three metrics between G and rest of the networks in the population, which follows $\mathbb{P}_{\mathfrak{D}_G}(A^*)$ and check if these distributions are similar. In the context of Figures 4-8 in the main text, we test if the distributions of black dots are robust to the choice of G^*.

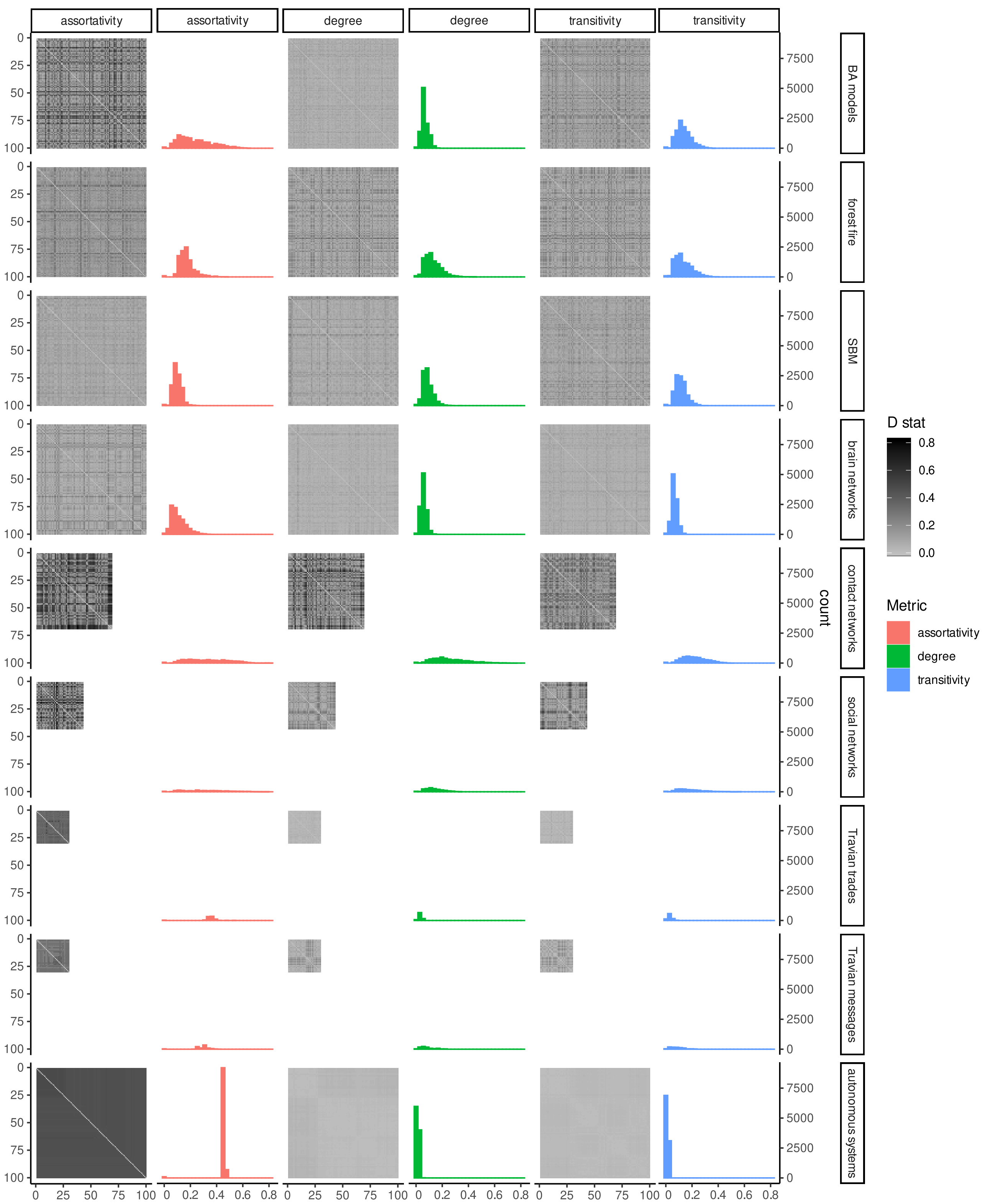

Figure 1: The pairwise KS statistic D and histogram of local assortativity, degree distribution and local transitivity on 9 network populations. 9 rows represent 9 sets of network population as labeled on the right. Columns 1, 3, and 5 represent the D statistic matrices of local assortativity, degree distribution and transitivity, respectively. X and Y axes in these columns indicate the number of networks in population. Columns 2, 4, and 6 show the histogram of the D statistic for all possible pairs of networks.

Figure 1 shows the KS statistic matrices $D(m)$ of the three network metrics m on the 9 network populations. Element $D_{ij}(m)$ represents the KS statistic of metric m between G_i and G_j. Based on the pairwise D statistics, we test if the distributions are similar for different G^*, i.e. if the rows in $D(m)$ are similar to each other.

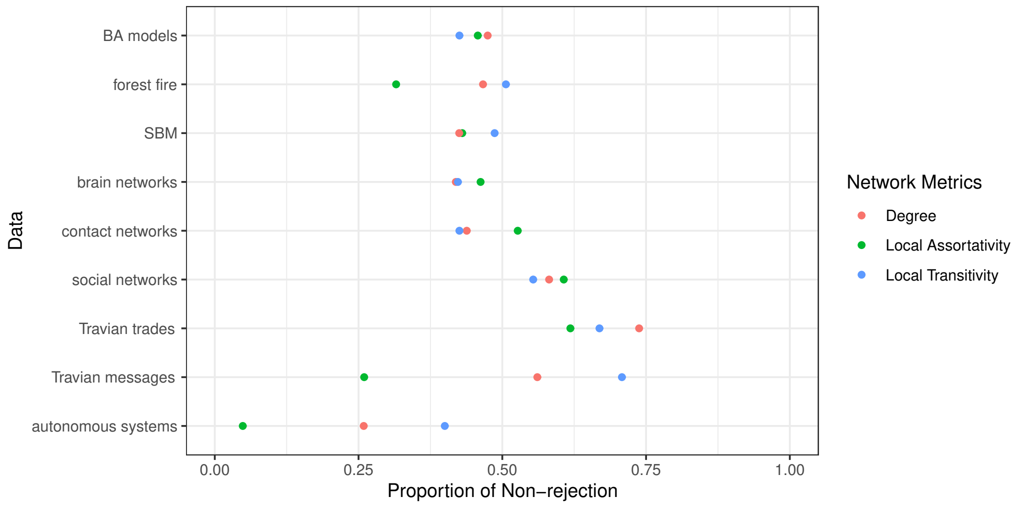

Figure 2: Proportion of non-rejection in KS test with 95% confidence level. Higher value indicates more pairs of G^* have same distribution.

Figure 2 shows the proportion of all tests that fail to reject the null hypothesis that the two samples are drawn from the same distribution. For most populations except autonomous systems, about half of the comparisons show no difference in dissimilarity distributions. In order to test this hypothesis together for all three metrics, multivariate (3 variables) two sample tests are performed with a kernel maximum mean discrepancy (kMMD) test. Similar to the KS test, the null hypothesis is that the two samples are drawn from the same distribution.

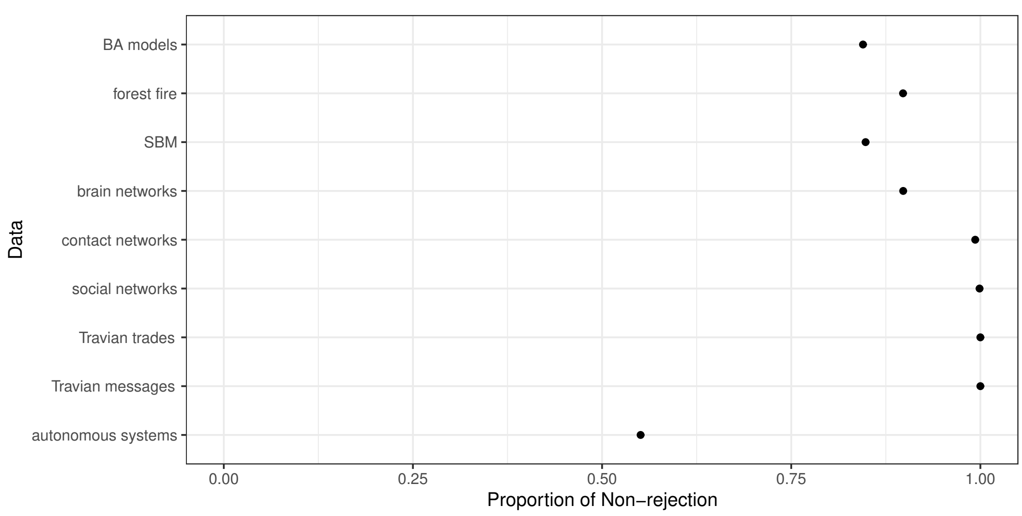

Figure 3: Proportion of non-rejection in kMMD test with 95% confidence level. Higher value indicates more pairs of G^* have same distribution.

Figure 3 shows the proportion of non-rejection in kMMD test. All populations have a non-rejection rate higher than 85% except autonomous systems. However, rejection in KS test or kMMD test does not revoke the result shown in Section 4 because the difference caused by choice of G^* is often negligible compared to the difference between true process and populations generated from generative models. For instance, the pair of networks in autonomous systems, #7 and #28, have the largest MMD among all pairs.

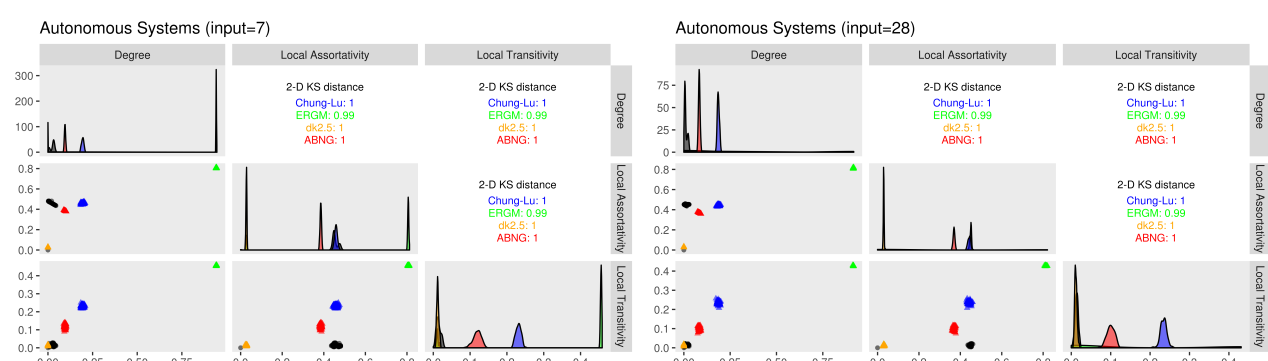

Figure 4: Empirical evaluation of the ability of network models to approximate the ground truth system based on observation of #7 network (left) and #28 network (right) in autonomous systems dataset.

The results shown in Figure 4 implies that the synthetic networks are consistent for different G^*, and the difference between ground truth and model simulation is much larger than the difference caused by different G^*.