Generating Networks to Target Assortativity via Archimedean Copula Graphons

Victory Idowu
victory.idowu.18@ucl.ac.uk

Department of Statistical Science
University College London, United Kingdom

May 2025

Motivation I

- In machine learning and network theory, we need to often create networks to meet a target level of assortativity, where assortativity is measured by the Newman assortativity coefficient.
- For instance, algorithms need to be tested on networks with predefined assortativity [9, 14] to conduct tasks like node classification [13] and link prediction [17, 1].
- Also, algorithms are required to efficiently operate on graphs with specified target motif counts [18, 12, 11] which are critical in tasks like motif discovery and graph matching [11].
- Current benchmark graph datasets often have predefined properties including assortativity and homomorphism densities [4, 3], but the process behind generation of these graphs are often opaque.

Motivation II

- However, random graph generators cannot generate graphs to a target level of assortativity without using rewiring.
- Rewiring mechanisms are often MCMC procedures that swap the connectivity of nodes until a criterion is met, the most common form of rewiring is that of Newman edge rewiring [10].
- Rewired graphs however may obtain target levels of assortativity at the expense of real world motif patterns [16, 7].
- The aim of this paper is to address the question:

Can we generate a random graph to a target level of assortativity, without having to rewire?

Paper link: `https://arxiv.org/abs/2503.03061`

Our Contribution

1. We propose a new graph generation algorithm only requiring a copula structure and number of nodes in the graphon framework.
2. We define the assortativity coefficient in terms of homomorphism densities of a graph or graphon
3. We show how to generate graphons using both the copula distribution function, the copula density function, and their tensor product.
4. We propose an algorithm to generate a network to a target level of assortativity using the copula graphon or copula density graphon; and another algorithm for their tensor product.

Simple graphs

A graph $G = (V, E)$, also referred to as $G(n, m)$:

- $V = \{1, 2, \ldots, n\}$ is the set of nodes
- $E \subseteq V \times V$ is the set of edges
- $|V| = n$ is the size (number of nodes)
- $|E| = m$ is the number of edges.

A network can then be represented by an adjacency matrix, i.e. $A = (A_{ij})$, where:

$$A_{ij} = \begin{cases} 1, & \text{if } ij \in E \\ 0, & else. \end{cases} \tag{1}$$

G is a simple graph if A is symmetric $A_{ii} = 0$, there are no loops and $A_{ij} = 0$ or $1, i \neq j$, there are no multiple edges between nodes.

Homomorphism densities

The homomorphism density between simple graphs [5], for $V(F) < V(G)$:

$$t\left(F, G\right) := \frac{\hom(F, G)}{n^m} \tag{2}$$

as well as the isomorphisms, t_{inj} or $|F|$, follows:

$$t_{\mathrm{inj}}\left(F, G\right) := \frac{\mathrm{inj}(F, G)}{(n)_m} \tag{3}$$

where $(n)_m$ is the falling factorial.
Note,

- $t(F, G) := \mathbb{P}(F \subseteq G[m])$ and
- $t_{\mathrm{inj}} := \mathbb{P}(F \subseteq G'[m])$

where $G[m]$, $G'[m]$ is the subgraph induced by picking m nodes of G with and without replacement respectively.

- A graphon $W \in \mathcal{W}$ is an integrable function $W : [0,1]^2 \to [0,1]$.
- The probability of an edge between points u_i, u_j is $W(u_i, u_j)$. $\{u_i\}$ are a collection of latent random variables and are often modeled as continuous uniforms.
- Let $\mathbb{G} = \mathbb{G}(n, W)$ be the random graph generated by the graphon W on n nodes, also referred to a $W-$random graph.
- The homomorphism density of F_G for a given graphon is expressed as:

$$t(F, W) = \int \prod_{i \in V(F)} dx_i \prod_{ij \in E(F)} W(x_i, x_j) \tag{4}$$

Note that $t_{\mathrm{inj}}(F, W) = t(F, W)$ since all maps of i occur with probability 1.

Copulas I

- $C : [0,1]^2 \to [0,1]$ is a bivariate copula if for every $u, v \in [0,1]$,

$$
\begin{aligned}
C(0,u) =& C(u,0) = 0 \\
C(1,u) =& C(u,1) = u \\
C(u,v) &= C(v,u).
\end{aligned}
$$

- For any $u_1 \leq v_1, u_2 \leq v_2$,

$$C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0.$$

- Note that copulas are defined on the cdf of the random variable whereas the joint pdf is defined on the value itself.

- All copulas are bounded above and below by the Fréchet-Höffding bounds [8]:

Theorem

For every bivariate copula, $C(u_1, u_2)$ the upper and lower bounds are given by:

$$C^{-}(u_1, u_2) \leq C(u_1, u_2) \leq C^{+}(u_1, u_2) \tag{5}$$

where $C^{-} := \max\{u_1 + u_2 - 1, 0\}$ *and* $C^{+} := \min\{u_1, u_2\}$.

- The limits of the Fréchet-Höffding bounds are called the countermonotonic (maximum) copula C^{-} and the comonotonic (minimum) copula C^{+}.

- The most basic copula is the independence copula $C(u_1, u_2) = u_1 u_2$.

- We will denote the independence copula by Π.
- C^{-}, C^{+} in addition with Π are called the *fundamental copulas*.

Archimedean copulas

General form of an Archimedean copula:

$$C\left(u_1, u_2\right) = \varphi^{[-1]}\left(\varphi(u_1) + \varphi(u_2)\right) \qquad \text{for } u_i \in [0,1]$$

where φ and $\varphi^{[-1]}$ are defined are generator and the (pseudo) generator inverse.

The generator function

A copula generator, φ, satisfies the following properties [6]:

1. $\varphi : [0, 1] \rightarrow [0, \infty]$
2. φ is a continuous and strictly decreasing function in the interval $[0, 1]$
3. φ is such that $\varphi(1) = 0$

The pseudo-inverse of φ within the domain $[0, \infty]$ is defined to be:

$$\varphi^{[-1]}(x) = \begin{cases} \varphi^{-1}(x) & 0 \leq x \leq \varphi(0) \\ 0 & \varphi(0) < x \leq \infty \end{cases} . \tag{6}$$

Gumbel copula I

Gumbel copula is used for dependence structures that show upper tail dependence [8]:

- The generator function: $\phi_G(t) = (-\log(t))^{\theta}$.
- The copula function:

$$C_G(u_1, u_2) = \exp\left(-\left((-\ln u_1)^{\theta} + (-\ln u_2)^{\theta}\right)^{\frac{1}{\theta}}\right) \tag{7}$$

for any $\theta \in [1, \infty)$.

Gumbel copula II

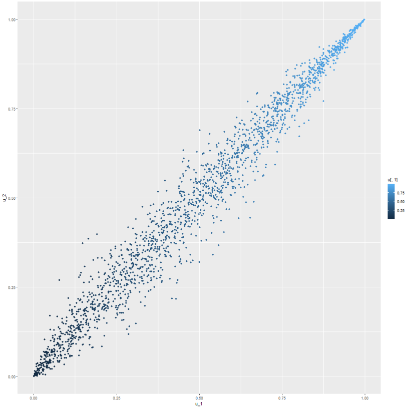

Figure 1: Gumbel copula $\theta = 10$

Clayton copula I

- The generator function: $\phi_C(t) = \frac{1}{\theta}(t^{-\theta} - 1)$.
- The corresponding Clayton copula is [8]:

$$C_C(u_1, u_2) = \{\max(u_1^{-\theta} + u_2^{-\theta} - 1, 0)\}^{-\frac{1}{\theta}} \tag{8}$$

for any $\theta \in [-1, \infty)\backslash\{0\}$.

Clayton copula II

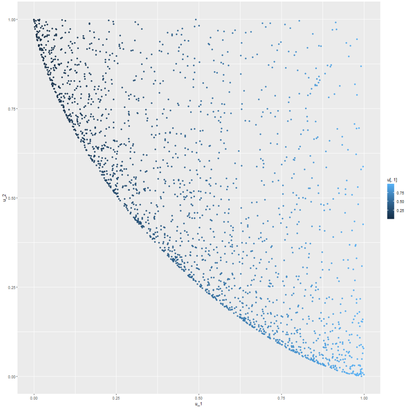

Figure 2: Clayton copula $\theta = -0.7$

Assortativity Coefficient I

- The (degree) assortativity coefficient is a graph parameter r_S, it is the Pearson correlation coefficient applied to the pairs of excess degrees on each edge of a network [9, 10]:

$$r_S = \frac{1}{\sigma_q^2} \sum_{jk} (e_{ij} - q_j q_k)$$

where q_i is the normalized excess degree distribution, e_{jk} is the joint excess degree probability distribution and σ_q^2 the variance of this distribution. $-1 \leq r_S \leq 1$.

Assortativity Coefficient II

- The empirical version of r:

$$r = \frac{m^{-1}\sum_i j_i k_i - \left[m^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2}{m^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - \left[m^{-1}\sum_i \frac{1}{2}(j_i + k_i)\right]^2} \tag{9}$$

where j_i, k_i are the degrees of vertices at the ends of the ith edge for $G = G(n, m)$.

Assortativity Coefficient III

- The formulation of the sample degree assortativity coefficient as a combinatorial ratio of isomorphism counts for a given graph G [2, 15]:

$$r = \frac{|P_2|(|P_{3/2}| + C - |P_{2/1}|)}{3|S_3| - |P_2|(|P_{2/1}| - 1)} \qquad (10)$$

  where P_i is the path on i nodes with $i-1$, S_i star graphs on i nodes C_i cycle graph on i nodes and $i+1$ edges, $C = \frac{3|C_3|}{|P_2|}$ is the clustering coefficient and $|P_{r/s}| := |P_r|/|P_s|$.

Assortativity Coefficient IV

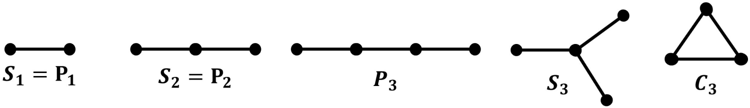

Figure 3: Star and path networks in the combinatorial expression of Newman's assortativity coefficient

By using (4), we can relate the degree-degree assortativity with the graphon function.

Theorem (Degree Assortativity Coefficient with Graphon Homomorphism Densities)

Let $W \in \mathcal{W}$ be a graphon on n nodes. The degree-degree assortativity coefficient characterised by homomorphism counts is:

$$r_W = \frac{(n-3)t(P_3, W) + \frac{3nt(C_3,W)}{(n-1)(n-2)} - \frac{(n-2)t(P_2,W)^2}{t(P_1,W)}}{(n-3)3t(S_3, W) + t(P_2, W) - \frac{(n-2)t(P_2,W)^2}{t(P_1,W)}}. \quad (11)$$

New Results: Assortativity Theorems II

If the theoretical homomorphism densities from a simple graph G is known, its degree assortativity is as follows.

Corollary

For a simple graph G, its degree assortativity coefficient is:

$$r_G = \frac{(n-3)t_{\mathrm{inj}}(P_3, G) + \frac{3nt(C_3,G)}{(n-1)(n-2)} - \frac{(n-2)t_{\mathrm{inj}}(P_2,G)^2}{t_{\mathrm{inj}}(P_1,G)}}{(n-3)3t_{\mathrm{inj}}(S_3, G) + t_{\mathrm{inj}}(P_2, G) - \frac{(n-2)t_{\mathrm{inj}}(P_2,G)^2}{t_{\mathrm{inj}}(P_1,G)}} \tag{12}$$

Indeed, this is immediate as that for graph F and simple G, $t_{\mathrm{inj}}(F, G) \neq t(F, G)$.

Theorem (Convergence to the Homomorphism Density Degree Assortativity Coefficient)

Let $r_{\mathbb{G}}$ be the assortativity coefficient of $\mathbb{G}(n, W)$ As $n \to \infty$, $r_{\mathbb{G}} \to r_W$.

Copula Graphons I

- Let G be a simple graph with graphon W with assumed copula structure C. The adjacency matrix A of G is as follows:

$$A_{ij}|\mathbf{u} \sim Bernoulli\left(C(u_i, u_j)\right) \tag{13}$$

  where u_i, u_j are uniform i.i.d random variables.

- Copula graphons generated from the four Archimedean copulas: Clayton, Frank, Gumbel and Joe are referred to as W_C, W_F, W_G and W_J respectively.

- We can express several graph statistics of the copula graphon in terms of the generator function.

Theorem (Edge Density of the Copula Graphon)

Under the copula graphon framework for dense graphs, the edge density, $t(P_1, W)$ is:

$$\frac{1}{2} - \int_1^{\varphi^{-1}(1)} \int_1^{\varphi^{-1}(1)} \varphi(x)\, \varphi'(x+y)\, \varphi'(y)\, dx\, dy. \tag{14}$$

Theorem (Degree operator of the Copula Graphon)

Under the copula graphon framework for dense graphs, the edge density is degree operator becomes:

$$\lambda(x) = -\int_x^\infty \frac{\varphi^{-1}(s)}{\varphi'(\varphi^{-1}(s-x))} ds \tag{15}$$

where $x = \varphi(u_1)$*. Moreover,* $\lambda(x)$ *is bounded above by* $\varphi^{-1}(x)$*.*

Theorem (Average degree of Archimedean copula graphons)

Let G be a simple graph with graphon W with assumed copula structure C. Then,

$$\frac{(n-1)}{6} \leq \mathbb{E}(d_i) \leq \frac{(n-1)}{3} \tag{16}$$

Theorem (Star Density)

The star density is $t(S_k, W)$ is:

$$\int_0^1 \lambda(x)^k dx = -\int_0^1 \left(\int_x^\infty \frac{\varphi^{-1}(s)}{\varphi'(\varphi^{-1}(s-x))} ds \right)^k dx \tag{17}$$

Copula Density Graphon I

The copula density graphon has a similar generative model to the copula graphon. Let G be a simple graph with graphon $\tilde{W}$ with assumed copula density function c. The adjacency matrix A of G is as follows:

$$A_{ij}|\mathbf{u} \sim Bernoulli\left(c(u_i, u_j)\right) \tag{18}$$

where u_i, u_j are uniform i.i.d random variables.

Theorem (Degree operator of the Copula Density Graphon)

For a copula density graphon $\tilde{W}$ with generator function φ, its degree operator has closed form:

$$\lambda(x) = \frac{\varphi'(\varphi^{[-1]}(x) + \varphi^{[-1]}(1))}{\varphi'(\varphi^{[-1]}(1))} - \frac{\varphi'(\varphi^{[-1]}(x) + 1)}{\varphi'(1)} \tag{19}$$

Theorem (Star density of Copula Density Graphon)

For a copula density graphon $\tilde{W}$, $t(S_k, W)$ is:

$$\int_0^1 \left(\frac{\varphi'(\varphi^{[-1]}(x) + \varphi^{[-1]}(1))}{\varphi'(\varphi^{[-1]}(1))} - \frac{\varphi'(\varphi^{[-1]}(x) + 1)}{\varphi'(1)} \right)^k dx \qquad (20)$$

Algorithm 1: Generative Pseudo Algorithm for a copula graphon or copula density graphon to desired r

Input: Function $F^\star$ which is bivariate Archimedean copula function C or bivariate Archimedean copula density function c; parameter θ; n number of nodes

Init: Generate an empty adjacency matrix A with n rows and n columns

for each $i = 1$ to j **do**

Generate A_{ij} which satisfies the following:

$$\mathbb{P}(A_{ij} = 1 | u_1, u_2) = F^\star(u_i, u_j) \tag{21}$$

Replace lower triangular matrix with transpose of upper triangular matrix.

Set diagonal= 0 **end for** set matrix $A = (A_{ij})$

Return: Simple graph W_G generated from A.

Algorithm 2: Generative Pseudo Algorithm of the tensor copula graphon

Input: A mixture $\mathbb{W}$ of bivariate Archimedean copula functions $\{C_{s_1}\}$ and / or copula density functions $\{c_{s_2}\}$ $s = s_1 + s_2$, $s > 1$, parameters $\{\theta_j\}_{j=1}^{s}$, n number of nodes

Init: Generate an empty adjacency matrix A with n rows and n columns
for each $i = 1$ to j **do**
Generate $2s$ i.i.d $u_i \sim U(0, 1)$
Generate A_{ij} which satisfies the following:

$$\mathbb{P}(A_{ij} = 1 | u_1, \dots, u_{2s}) = \prod_{t=1}^{s} \hat{W}_t(u_{2i-1}, u_{2j}) \tag{22}$$

Replace lower triangular matrix with transpose of upper triangular matrix.
Set diagonal= 0 **end for** set matrix $A = (A_{ij})$

Return: Simple graph W_G generated from A.

Numerical Examples I

The configuration model and preferential attachment produces random graphs with $r = 0$. Likewise, this can also be achieved with the copula graphon by using Π or copula graphons that will approximate Π for some θ in their parameter space. Note that Π does not induce any dependency structure between the nodes. For $W \sim \Pi$, $C \approx 0.04$ and thus $|P_{3/2}| \approx |P_{2/1}|$.

Numerical Examples II

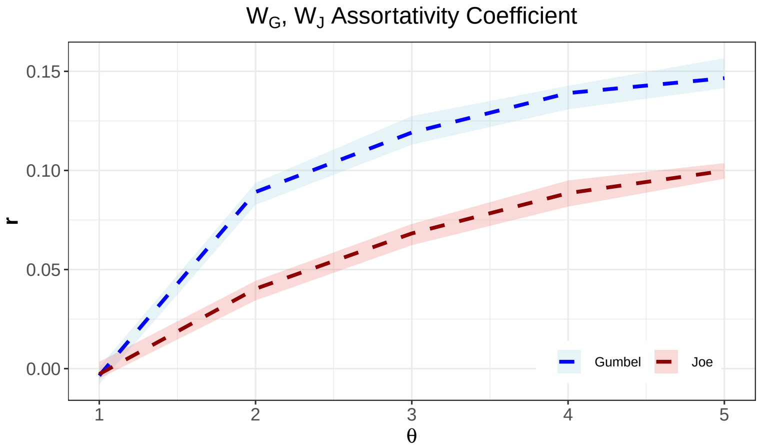

Figure 4: Assortativity coefficient range by W_G, W_J for $\theta \in [0, 10]$. Blue, red line is the average assortativity coefficient of W_G, W_J respectively, over 10 repetitions, the minimum and maximum within the range is the shaded region

Numerical Examples III

Assortative networks have stronger mixing patterns with nodes of higher degree [9], which cannot be achieved by the copula graphons alone.

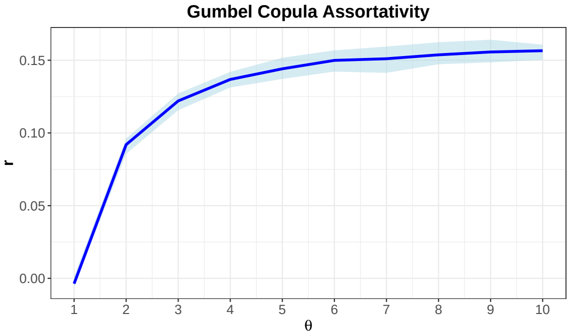

Figure 5: Assortativity coefficient range by W_G for $\theta \in [0, 10]$. Blue line is the average assortativity coefficient over 10 repetitions, the minimum and maximum within the range is the shaded region

Numerical Examples IV

Disassortative networks can be generated by tensor product of weakly disassortative copula graphons. The tensor product can create more intricate/pronounced mixing patterns. The extent of disassortativity can be controlled by lagging θ.

Numerical Examples V

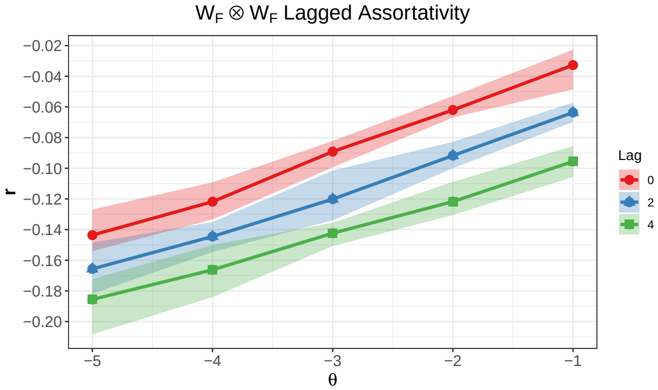

Figure 6: Assortativity coefficient range through $\tilde{W}_F \otimes \tilde{W}_F$ for $\theta_F \in [1, 5]$ at different lags. Assortativity coefficient is averaged over 10 repetitions. Line is the average assortativity coefficient over 10 repetitions, the minimum and maximum within the range is the shaded region

Conclusion and discussion

- We have developed a method using copulas and graphons to generate a network to meet a target level of assortativity without using rewiring.
- We conducted extensive experiments to explore the possible assortativity different within the parameter space of selected copula graphons, copula density graphons and their tensor products.
- In the future, we will explore the possibility of using non-Archimedean copula structures to generate networks to target assortativity.

Relevant References I

[1] Eduardo G Altmann and Alexey Vlaskin. "Synthetic graphs for link prediction benchmarking". In: *Journal of Physics: Complexity* (2024).

[2] Ernesto Estrada. "Combinatorial study of degree assortativity in networks". In: *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 84 (2011).

[3] Weihua Hu et al. "Open graph benchmark: Datasets for machine learning on graphs". In: *Advances in neural information processing systems* 33 (2020).

[4] Renming Liu et al. "Taxonomy of benchmarks in graph representation learning". In: *Learning on Graphs Conference*. PMLR. 2022.

Relevant References II

[5] László Lovász and Balázs Szegedy. "Limits of dense graph sequences". In: *Journal of Combinatorial Theory, Series B* 96 (2006).

[6] Alexander J McNeil, Rüdiger Frey, and Paul Embrechts. *Quantitative risk management: concepts, techniques and tools-revised edition*. Princeton university press, 2015.

[7] David Melamed, Ashley Harrell, and Brent Simpson. "Cooperation, clustering, and assortative mixing in dynamic networks". In: *Proceedings of the National Academy of Sciences* 115 (2018).

[8] Roger B Nelsen. *An introduction to copulas*. Springer, 2006.

Relevant References III

[9] Mark EJ Newman. “Assortative mixing in networks”. In: *Physical review letters* 89 (2002).

[10] Mark EJ Newman. “Mixing patterns in networks”. In: *Physical review E* 67 (2003).

[11] Pedro Ribeiro et al. “A survey on subgraph counting: concepts, algorithms, and applications to network motifs and graphlets”. In: *ACM Computing Surveys (CSUR)* 54 (2021).

[12] Ryan A Rossi, Rong Zhou, and Nesreen K Ahmed. “Estimation of graphlet counts in massive networks”. In: *IEEE transactions on neural networks and learning systems* 30 (2018).

Relevant References IV

[13] Susheel Suresh et al. “Breaking the limit of graph neural networks by improving the assortativity of graphs with local mixing patterns”. In: *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*. 2021.

[14] Anton Tsitsulin et al. “Synthetic graph generation to benchmark graph learning”. In: *arXiv preprint arXiv:2204.01376* (2022).

[15] Demival Vasques Filho and Dion RJ O’Neale. “Transitivity and degree assortativity explained: The bipartite structure of social networks”. In: *Physical Review E* 101 (2020).

[16] S Xiao et al. “Robustness of scale-free networks under rewiring operations”. In: *Europhysics Letters* 89 (2010).

Relevant References V

[17] Xiao-Ke Xu, Ke-Ke Shang, and Jing Xiao. "Quantifying the effect of community structures for link prediction by constructing null models". In: *IEEE Access* 8 (2020).

[18] Zhao Zhao. "Subgraph querying in relational networks: a mapreduce approach". In: *2012 IEEE 26th International Parallel and Distributed Processing Symposium Workshops & PhD Forum*. IEEE. 2012.