

Limit Theorems with Multiple Small World Clusters/Networks*

Arkadiusz Szydlowski[†]
University of Kent

October 23, 2026

Abstract

We refine the conditions needed to obtain central limit theorems for data with multiple clusters or networks allowing for different patterns of growth rates in the size and the number of clusters/networks, which correspond to different network sampling designs or network evolution schemes. We consider both the means of node- and edge-specific characteristics. Next, we specialise the results to the case of sparse, small world networks, i.e. networks with limited maximal degree and a relatively small (but growing) diameter, properties encountered in many social and economic networks. In this setting, the conditions translate into restrictions on the constant of growth of the network diameter relative to the maximal degree. Our Monte Carlo simulations confirm that the wild cluster bootstrap provides accurate inference when the subnetworks are homogenous but undercovers otherwise.

JEL: C10, C45

Keywords: Networks, Clusters, CLT, Wild cluster bootstrap

1 Introduction

In many cases economic data come from several groups or individuals interconnected through their social networks and it is important to adjust inference for the presence of related dependence between the observations. In particular, it is important to know conditions

*I would like to thank the editors and two anonymous referees for very useful comments that greatly improved the article. I have also benefitted from feedback provided by the Bristol Econometric Study Group participants, Vadim Marmer, Nizar Allouch and Alfred Duncan. For the purpose of open access, the author(s) has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising.

[†]School of Economics, Politics and IR, University of Kent, LASE Group, Canterbury, CT2 7FS, UK.
E-mail address: a.szydlowski@kent.ac.uk

under which asymptotic theory can be used to approximate distributions of the sample statistics. When the number of clusters stays the same once we increase the sample size, such conditions have been provided by Bester et al. (2011) and then extended by Djogbenou et al. (2019) (DMN, henceforth) and Hansen & Lee (2019) to a growing number of clusters.

As our first contribution, we refine the latter results by distinguishing cases when we may have various proportions of growing and fixed-size clusters or network components and show that the restrictions on the size of the largest cluster in DMN can be relaxed in some cases. This refinement is motivated by empirical studies in which network evolution involves components growing at different rates alongside an increasing number of fixed-size components, as well as different network sampling schemes (cf. Kolaczyk (2009), Ch. 5.3). For example, Tomassini & Luthi (2007) demonstrate that the coauthorship network in genetic programming evolved in that fashion, with increasing share of nodes in the giant component, decreasing share of nodes (but increasing number) in the second largest component, and the number of components growing with the number of nodes.

The conditions needed to obtain a CLT vary across different network evolution structures, in particular with respect to the number of growing components and the variation of their sizes. At one end, when the network consist of many components growing at the same rate it is enough that the largest component grows at a rate only marginally slower than N , at the other end, when we have both large components growing in size and many fixed-size components, the largest component may need to grow at a rate smaller than $\sqrt{N}$.

As a second contribution, we apply these results to provide low level conditions for the case of sparse, small world networks. Many economic and social networks on top of being sparse exhibit a small-world property (Watts & Strogatz (1998)), namely that the network distance between each pair of connected nodes is small compared to the number of nodes in the network.¹ Formally, a small-world network has a diameter proportional to $\log N$, where N is the number of nodes. We combine this restriction and sparsity and show that a CLT applies to data coming from multiple small world networks under additional assumptions restricting the size of the diameter relative to the maximal degree for any given N , or, put differently, restricting the constant of proportionality relating the diameter to $\log N$.

As a third contribution, in addition to node-specific means we also consider CLTs for edge-specific characteristics. We show that a CLT for the means of edge characteristics holds under a relatively mild and natural strengthening of conditions needed for node-specific means. The result applies to network U-statistics under many subnetworks asymptotics. As the clustering literature rarely imposes network structure and, hence, does not define edge-specific means, our results seem novel here.

¹The “small-world” property often also includes the characteristic that the network graph is much more clustered than a random graph (see e.g. Definition 4.1.3 in Watts (1999)). However, the latter property is not useful for the purpose of providing CLT in our context so we do not discuss it.

Overall, our contribution compared to DMN amounts to relaxing some of the cluster-size restrictions for node-means CLT due to a more careful bounding argument. Our restrictions on the size of the largest cluster are also sometimes weaker than the ones in Hansen & Lee (2019), though our assumptions are not directly comparable as they utilise Lindeberg-Feller’s approach to CLT and require uniform integrability. In general, we provide lower-level conditions than these papers and link them to the strength of within-cluster dependence. An important distinguishing feature of this article is also that we propose completely new set of conditions for a CLT for pair-level statistics.

Kojevnikov et al. (2021) provide a CLT for node-specific means for data coming from one large network assuming weak dependence between the nodes in the form of ψ -dependence (see also Leung & Moon (2026)) but they do not analyse edge-specific means. Ogburn et al. (2024) provide a CLT for network data with unrestricted correlation within neighbourhoods but only allow dependence up to friends-of-friends, whereas we allow for any connected nodes (via any path) to be dependent.

There is a large literature on obtaining limit theorems with spatial networks (see Jenish & Prucha (2012), Kuersteiner & Prucha (2013), Kuersteiner (2019) among others). Although many social and economic phenomena can be modelled using these networks, most social and economic networks feature cliques (see Jackson (2008)), which spatial networks have trouble accommodating, particularly when the maximal clique size grows with N (Kojevnikov et al. (2021)).

Technically, our results rely on verification of Lyapunov’s condition, just as DMN. They can also be seen as related to the results on limit theorems with m -dependence, where m can diverge to infinity (Romano & Wolf (2000)), and the literature on normal approximations under local dependence (Baldi & Rinott (1989), Chen & Shao (2004)). Other recent articles on inference using network data include Bickel et al. (2011), Bickel et al. (2013), Matsushita & Otsu (2024) among others.

As we aim to apply our results to small world networks, we refer to different dependency groups in the data as network components, rather than clusters. All proofs are relegated to the online supplement.

2 Main idea

Let $\{Y_1, \dots, Y_N\}$ denote random variables corresponding to nodes in a network G_N . We are interested in the central limit theorem for the sample mean:

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

assuming that only nodes not connected through any network path have statistically independent Y_i 's and dependence between the remaining observations is not restricted. For example, inference on the network sample average is of interest in the studies of economic connectedness (Chetty et al. (2022)): we may want to test if we have perfectly economically connected communities by testing $H_0 : E[Y_i] = 0.5$, where Y_i is the individual economic connectedness for a person i in the low SES (Socio-Economic Status) group, defined as a fraction of friends in the high SES group among all friends.

Let the network consist of c_N separated components of size $N_c, c = 1, \dots, c_N$, and the number of non-zero pairwise correlations between the Y_i 's in a component be of order $N_c^{1+\gamma_c}, \gamma_c \in [0, 1]$. Also, let $c = 1$ correspond to the component for which $N_c^{1+\gamma_c}$ grows the fastest. Then, if $0 < \text{Var}(Y_i) < \infty$, we have:

$$\text{Var} \left(\sum_{i=1}^N Y_i \right) \sim \sum_{c=1}^{c_N} N_c^{1+\gamma_c}$$

Using this structure, we provide sufficient conditions for the Lyapunov condition that involve bounds on the growth rates of N_c 's. Next, the bound on the maximal number of nodes in a sparse network with a given maximal degree and diameter (Pineda-Villavicencio & Wood (2015)) is used to translate these conditions to the parameters of a small world network.

The refinement compared to DMN comes from the fact that they consider a general case with all clusters possibly growing with N . Thus, in order to verify the Lyapunov's condition, they bound:

$$N^{-2-\delta} \sum_{c=1}^{c_N} N_c^{(1+\gamma_c)(2+\delta)} \leq N^{-1-\delta} N_1^{(1+\gamma_1)(1+\delta)}$$

and the right-hand side converges to zero if $N_1^{1+\gamma_1}/N = \eta_N^{-1/(1+\delta)}$ where $\eta_N \rightarrow \infty$. However, if the network evolves in such a way that the number of growing components (c_k) stays the same and the number of components of fixed size ($c_N - c_k$) grows with N (see Thm. 2(a)), we can bound:

$$N^{-2-\delta} \sum_{c=1}^{c_N} N_c^{(1+\gamma_c)(2+\delta)} \leq c_k N^{-2-\delta} N_1^{(1+\gamma_1)(2+\delta)} + O(N^{-1-\delta})$$

and now the right-hand side will converge to zero if $N_1^{1+\gamma_1}/N = \eta_N^{-1/(2+\delta)}$. This accommodates "larger" components N_1 than the previous condition, thus, allows us to relax the assumptions in DMN in this case. Similar refinements follow under other network evolution assumptions.

Since our conditions imply N_c is of order smaller than N , the results do not apply to

networks with a giant component involving almost all nodes or, alternatively, networks with a giant component growing at rate N . This happens, for example, in the Facebook network where 99.91% of individuals belong to the largest connected component (Ugander et al. (2011)). However, if one is interested in actual dependence between individuals on Facebook, instead of using formal links, one may only keep connections that generate some cross-traffic (views, shares, likes etc.), thus splitting the giant component into multiple groups that can be viewed as independent, which would fit into our framework.

3 Central limit theorems for small-world networks

For simplicity, assume that Y is centred, i.e. $E[Y] = 0$. Let $d_{max} = \max_{i \in [N]} d_i$ be the maximal degree in network G_N . Define $l(i, j)$ to be the network distance on the shortest path between i and j and set $l(i, j) = \infty$ if i and j are not connected by any network path. The diameter of network G_N is now formally defined as $\Delta_N = \max_{i, j \in [N]: l(i, j) < \infty} l(i, j)$. All our results hold conditionally on network evolution $\{G_N\}_{N=1}^\infty$.

3.1 Node specific means

Consider the sample mean $\bar{Y}$ defined above, let $C_c \subset [N]$ enumerate nodes in a network component c and make the following assumptions:

Assumption CVAR. (a) $l(i, j) = \infty$ implies $Y_i \perp Y_j$.

(b) There exist $\{\gamma_c\}_{c=1}^{c_N} : 0 \leq \gamma_c \leq 1, \underline{\sigma}^2 > 0, \bar{\sigma}^2 > 0, \delta > 0$ such that:

$$\underline{\sigma}^2 N_c^{1+\gamma_c} \leq \text{Var} \left(\sum_{i \in C_c} Y_i \right) \leq \bar{\sigma}^2 N_c^{1+\gamma_c}, \quad (1)$$

$$E \left(\sum_{i \in C_c} Y_i \right)^{2+\delta} \leq K_N^{\frac{2+\delta}{2}} N_c^{(1+\gamma_c)\frac{2+\delta}{2}}. \quad (2)$$

for all c and N , where $K_N = O(1)$.

Part (a) states that unconnected nodes are statistically independent. The second condition has two parts. The first part assumes that the number of non-zero correlations in a component of size N_c is a power function of N_c . Though, certainly imposing some structure on the number of non-zero correlations, it accommodates a variety of cases, in particular, all correlations being non-zero ($\gamma_c = 1, \underline{\sigma}^2 < 1, \bar{\sigma}^2 > 1$), the fraction of non-zero correlations being constant ($\gamma_c = 1, \underline{\sigma}^2 < 1, \bar{\sigma}^2 < 1$) and decreasing toward zero with N_c ($\gamma_c < 1$). We note that the power form is chosen for convenience and clarity of the CLT conditions and

one may impose other correlation structures and follow similar arguments as in our proofs to obtain alternative conditions.² Also, one does not need to know γ_c 's in practice and can assume $\gamma_c = 1$, for all c , for the purpose of the conditions given below. Similarly, Romano & Wolf (2000) impose conditions that imply $Var(\sum_{i=1}^N Y_i) \sim N^{1+\gamma}$ for some $-1 \leq \gamma < 1$.

The second part of (b) assumes existence of moments for the component-wise averages. The assumption encompasses many dependence cases, including fully correlated outcomes and weak dependence. Without restrictions on intra-component dependence, we can bound using Minkowski's inequality: $E(\sum_{i \in C_c} Y_i)^{2+\delta} \leq O(N_c^{2+\delta} \sup_i E|Y_i|^{2+\delta})$, so a sufficient condition for (2) is that Y_i has $2 + \delta$ moments. When Y_i is a martingale difference sequence, Theorem 2.12 in Hall & Heyde (1980) implies that $E(\sum_{i \in C_c} Y_i)^{2+\delta} \leq O(\max\{E[(\sum_{i \in C_c} E[Y_i^2 | \mathcal{F}_{i-1}])^{(2+\delta)/2}], N_c \sup_{i \in C_c} E|Y_i|^{2+\delta}\})$, thus, if both $E[Y_i^2 | \mathcal{F}_{i-1}]$ and $E|Y_i|^{2+\delta}$ are bounded, we can write $E(\sum_{i \in C_c} Y_i)^{2+\delta} \leq O(N_c^{(2+\delta)/2})$ and condition (2) is satisfied with $\gamma_c = 0$. The latter bound can also be obtained for absolutely regular processes under summability conditions on the β -mixing coefficients (see e.g. Theorem 2.2. in Viennet (1997)).

Further conditions needed to obtain a central limit theorem depend on the network evolution scheme, in particular, how many growing network components there are and what their relative growth rates are. Before we proceed, we need to clarify some notation. As the network evolves both the existing components grow in size and new components arise. Thus, we will make the dependence of the size of a component on the number of nodes explicit, i.e. write $N_c(N)$, and keep in mind that c_N is a function of N . Without loss of generality, let the first component ($c = 1$) be a component for which the number of non-zero correlations, i.e. $N_1^{1+\gamma_1}$, grows at the fastest rate.

3.1.1 Many growing components

We start with a case of many growing components, where the components may grow at the same or different rates. Practically, this case corresponds to school networks as in Paluck et al. (2016) – increasing the sample size means adding more schools and more pupils to schools. Here we may have either the network component (school) sizes evolving similarly or some components growing faster than others (e.g. large schools in large cities may be oversampled).³

Alternatively, this case can be linked to the iterated star sampling, where we first sample a group of nodes and their immediate links, in the next round we sample links of the linked nodes etc. If the network components are large that will results in all components

²As $N_c^{1+\gamma_c}$ does not need to be an integer, the bounds $\sigma^2, \bar{\sigma}^2$ also accommodate rounding.

³In general, one may think of increasing N as sampling more nodes from an existing network **or** a fully observed network dynamically evolving as we allow dependence between all connected nodes.

growing when we increase N . Similarly, this will happen with link-tracing sampling where in each sampling round we map out network paths between randomly drawn source and target nodes, and induced and incident subgraph sampling where only links/nodes between randomly selected nodes/links are recorded in each sampling round (see Kolaczyk (2009) for more details).⁴

Theorem 1. *Let $\{Y_i\}_{i=1}^\infty$ be a sequence of mean zero random variables and define $B_N^2 = \text{Var}(\sqrt{NY})$. Under Assumption CVAR:*

$$\frac{\sqrt{NY}}{B_N} \rightarrow^D N(0, 1)$$

as $N \rightarrow \infty$ (conditionally on network evolution) if either of the following holds:

- (a) all components grow at the same rate, γ_c 's are all equal and $N_1/N \rightarrow 0$,
- (b) all components grow (possibly at different rates) and we have for all c :
 - (i) the component with the fastest growing number of intra-component correlations is also the fastest growing component, i.e. $\frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0$ implies $\frac{N_c(N)}{N_1(N)} \rightarrow 0$,
 - (ii) $\frac{N_c(N)^{\gamma_c}}{N_1(N)^{\gamma_1}}$ is weakly decreasing in N ,
 - (iii) there exists $M < \infty$ such that $\frac{N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}}}{N_{c_N}(N)^{\gamma_N}} < \left(\frac{N_1(\tilde{N})}{N_1(N)}\right)^{\gamma_1}$ for all $(N, \tilde{N})$ such that $c_{\tilde{N}} = c_N + 1$ and $N > M, \tilde{N} > M$,
 - (iv) $N_1^{(1+\gamma_1)\frac{2+\delta}{\delta}}/N \rightarrow 0$.

The results can be restated as results conditional on common shocks affecting all the nodes in the network just as in Kojevnikov et al. (2021), and would then apply to networks where there is some dependence between unconnected nodes and the dependence can be modelled through observables.

The condition in (a) prevents presence of a giant component of size $O(N)$ and requires the amount of dependence within the subnetworks to be similar, still allowing for subnetworks of different sizes. Conditions (ii) and (iii) assume that the largest component is always dominant in terms of the within-component correlations (as controlled by N_c and γ_c). In words, (ii) assumes that none of the existing components can "overtake" the largest component in that metric and (iii) means that any newly appearing component $(N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}})$ cannot be larger compared to the last appearing component $(N_{c_N}(N)^{\gamma_N})$ than the new size of the largest component $(N_1(\tilde{N})^{\gamma_1})$ relative to the last size $(N_1(N)^{\gamma_1})$. Last condition in part (b) becomes $N_1^{1+\gamma_1}/N \rightarrow 0$ if all moments of Y exist ($\delta = \infty$).

⁴Sampling may introduce bias into the estimated mean. Dealing with such bias is beyond the scope of this article.

In the school example condition (b) would imply that we do not oversample smaller schools, larger schools are sampled first and that at some point we only add to the sample similar small schools (while still sampling from the largest school). The condition (b) should also be naturally satisfied with star or link-tracing sampling as the nodes belonging to the largest component have a higher probability of being drawn into the first step of these sampling methods, and thus to be sampled quicker. In particular, link-tracing algorithms like TRACEROUTE (Kolaczyk (2009), Ch. 5.3.3) will not only draw nodes from the largest components with higher probability but they will also discover more nodes in the larger components in the path-tracing stage.

Remark 1. *The first condition in (b) requires that the component(s) with the fastest growing number of non-zero between-nodes correlations is(are) also the fastest growing component(s). It is trivially satisfied if all γ_c 's are equal. Additionally, note that together conditions (b)(i)-(ii) imply $\frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0 \Leftrightarrow \frac{N_c(N)}{N_1(N)} \rightarrow 0, \forall c > 1$.*

Remark 2. *DMN use Lyapunov's theorem, so serve as a natural reference point for our results. However, we may also compare our conditions to Hansen & Lee (2019) (HL19) who apply the Lindeberg-Feller theorem. Relationship between our and their conditions is subtle as they require uniform integrability of $|Y_i|^{2+\delta}$, which is not implied by our assumption CVAR (b). Thus, in general, neither set of conditions seems stronger or weaker and there is a tradeoff between approaching the problem from their and our perspective.*

Let $\gamma_c = 1$ for all c . For case (a) in Theorem 1 HL19 require $N_1^2/(N\lambda_N) \rightarrow 0$, where $\lambda_N \leq O(N_1)$, which amounts to our condition $N_1/N \rightarrow 0$ if λ_N achieves the upper bound. Given that uniform integrability implies the second part of CVAR (b), we can conclude that for this case our conditions are weaker when all outcomes within components are correlated (i.e. first part of CVAR (b) is satisfied).

For case (b) their main conditions require $N_1^2/N \rightarrow 0$ and $(\sum_{c=1}^{c_N} N_c^{2+\delta})^{2/(2+\delta)}/N < \infty$. Using $N_c \leq N_1$ and $c_N < N$ we can bound:

$$\frac{(\sum_{c=1}^{c_N} N_c^{2+\delta})^{\frac{2}{2+\delta}}}{N} < \left(\frac{N_1^{\frac{2(2+\delta)}{\delta}}}{N} \right)^{\frac{\delta}{2+\delta}},$$

so the condition for the right hand side to converge to 0 is the same as our condition (b)(iv). Here note that with $\gamma_c = 1$ the assumptions on the rate of growth of new components versus the largest component (b)(i)-(iii) are not needed (see Online Supplement B for more discussion).

3.1.2 Growing and fixed-size components

In some settings, not all network components grow with the sample size. For example, in the frequently used AddHealth data social networks of pupils in 16 schools were fully sampled by in-home interviews and the rest only partially-sampled where each pupil could name up to 5 friends from their school roster (see e.g. Harris et al. (2006)). Thus, if we wanted to increase the sample size from the existing network, we would only sample from the partially-sampled schools and new schools. So this would be an asymptotic scheme where some components have size that does not change with N and some grow with N .

If we interpret increasing N as a non-equilibrium network evolving dynamically, the cases in this section seem to also correspond well to the network evolution in genetic programming coauthorship network discussed in Tomassini & Luthi (2007), where there is a large growing group of widely collaborating researchers and many smaller groups whose members collaborate only among themselves.

Additionally, snowball sampling would often lead to a similar asymptotic scheme. Here we sample nodes and their friends and friends-of-friends etc. up to distance K from the original node. Thus, smaller components are likely to be fully sampled within the K steps of the original sampling round but larger ones would only be sampled in the subsequent sampling rounds. This would produce component size evolution in which some components have fixed size but some grow as N increases.

Theorem 2. *Let $\{Y_i\}_{i=1}^\infty$ be a sequence of mean zero random variables and define $B_N^2 = \text{Var}(\sqrt{NY})$. Under Assumption CVAR:*

$$\frac{\sqrt{NY}}{B_N} \rightarrow^D N(0, 1)$$

as $N \rightarrow \infty$ (conditionally on network evolution) if either of the following holds:

- (a) c_k components grow with N and remaining $c_N - c_k$ components have fixed size, c_k is fixed and $N_1^{1+\gamma_1}/N \rightarrow 0$,
- (b) c_k components grow with N and remaining $c_N - c_k$ components have fixed size, $\frac{c_N}{N} \rightarrow s > 0$, $c_k \rightarrow \infty$ and one of the following conditions is satisfied:
 - (i) $N_1^{(1+\gamma_1)\frac{2+\delta}{\delta}}/N \rightarrow 0$,
 - (ii) $\frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0$ implies $\frac{N_c(N)}{N_1(N)} \rightarrow 0$, $\frac{N_c(N)^{\gamma_c}}{N_1(N)^{\gamma_1}}$ is weakly decreasing in N , there exists $M < \infty$ such that $\frac{N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}}}{N_{c_N(N)}^{\gamma_N}} < \left(\frac{N_1(\tilde{N})}{N_1(N)}\right)^{\gamma_1}$ for all $(N, \tilde{N})$ such that $c_{\tilde{N}} = c_N + 1$ and $N > M, \tilde{N} > M$, and $N_c^{1+\gamma_c}/N \rightarrow 0, \forall c$,

(c) c_k components grow with N and remaining $c_N - c_k$ components have fixed size, $\frac{c_N}{N} \rightarrow 0$ and we have for all c :

- (i) $\frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0$ implies $\frac{N_c(N)}{N_1(N)} \rightarrow 0$,
- (ii) $\frac{N_c(N)^{\gamma_c}}{N_1(N)^{\gamma_1}}$ is weakly decreasing in N ,
- (iii) there exists $M < \infty$ such that $\frac{N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}}}{N_{c_N(N)}^{\gamma_N}} < \left(\frac{N_1(\tilde{N})}{N_1(N)}\right)^{\gamma_1}$ for all $(N, \tilde{N})$ such that $c_{\tilde{N}} = c_N + 1$ and $N > M, \tilde{N} > M$,
- (iv) $N_1^{(1+\gamma_1)\frac{2+\delta}{\delta}}/N \rightarrow 0$ when $\frac{c_k}{c_N} \rightarrow 1$ or $N_1^{\frac{2+\delta}{\delta}}/N \rightarrow 0$ otherwise.

Case (a) would correspond to AddHealth data when we would leave the number of fully-sampled schools fixed and only obtain additional data from partially-sampled and new schools. Assuming $\gamma_1 = 1$, the condition in this part implies that the largest component has to be of size smaller than $O(\sqrt{N})$.

Empirically, cases (b)-(c) would correspond to AddHealth data sampling in which we add fully-sampled schools in subsequent sampling rounds. They seem also particularly well suited to coauthorship network evolution as in Tomassini & Luthi (2007), where we observe more components when new researchers enter the network. The difference between case (b) and (c) lies mainly in how fast the number of components grows with N – proportionally or less than proportionally, and will usually be specific to the analysed network.⁵

The results in Theorem 1 and 2 refine the results in DMN by distinguishing different sampling/network evolution schemes. This often produces relaxed conditions on the permissible size of the largest cluster/network component. To see that, adapting from their regression setup to a sample mean setup, if η_N denotes the rate of divergence of $\text{Var}(\sum_{i=1}^N Y_i)$, they require (see MacKinnon et al. (2023)): $\left(\frac{\sqrt{\eta_N}}{N}\right)^{-(2+\delta)/(1+\delta)} \frac{N_1}{N} \rightarrow 0$. Let us compare this condition to the ones in part (a) of Theorem 1 and 2. For the former case, it is easy to derive that both our and their conditions require $c_N \rightarrow \infty$, in other words $N_1/N \rightarrow 0$. For the latter case, we need $N_1^2/N \rightarrow 0$ (assuming $\gamma_1 = 1$), whereas their condition implies that we need:⁶

$$\left(\frac{N^2}{\sum_{c=1}^{c_k} N_c^2 + (c_N - c_k)O(1)}\right)^{\frac{2+\delta}{2(1+\delta)}} \frac{N_1}{N} = \left(\frac{N^{-\frac{\delta}{2+\delta}} N_1^{\frac{2(1+\delta)}{2+\delta}}}{N^{-1} \sum_{c=1}^{c_k} N_c^2 + O(1)}\right)^{\frac{2+\delta}{2(1+\delta)}}$$

to converge to zero, which is satisfied if $N_1^{2(1+\delta)/\delta}/N \rightarrow 0$. Note that this is a stronger condition (e.g. requires $N_1^3/N \rightarrow 0$ when $\delta = 2$) than the one we impose and both are

⁵A quick look at Fig. 1 and Fig. 3 in Tomassini & Luthi (2007) suggests that $c_N/N \rightarrow 0$ in genetic programming coauthorship network so case (c) is relevant here.

⁶Note that the second part of our Assumption CVAR(b) (with $\gamma_c = 1$) is implied by an equivalent of their Assumption 1 (see e.g. their Lemma A.2).

equivalent only if all moments exist (i.e. $\delta = \infty$). Thus, in brief, our conditions allow larger components than DMN.⁷

Mathematically, DMN bound quantities like $\sum_{c=1}^{c_N} N_c^{1+\xi}$ by $c_N N_1^{1+\xi}$, which seems natural in their clustering setup with fairly homogenous components. We exploit the structure of different asymptotic regimes to provide a refined bounding argument, e.g. in case (a) of Theorem 2, we can bound $\sum_{c=1}^{c_N} N_c^{1+\xi} \leq c_k N_1^{1+\xi} + (c_N - c_k)O(1)$, which grows slower than the DMN bound and allows for weaker conditions on N_1 in order to obtain convergence.

Remark 3. *Note that the rate condition on N_1 in (b)(ii) is weaker than the one in (b)(i), but both become equivalent to conditions in (c)(iv) and (a) when $\delta \rightarrow \infty$.*

Remark 4. *Let us compare rate conditions in Theorem 2 to those in HL19. Again, for simplicity, let $\gamma_c = 1$ for all c . For case (a) in Theorem 2, HL19 conditions simplify to $N_1^2/N \rightarrow 0$ and this is exactly the same as our rate condition in (a). To see that note that $(\sum_{c=1}^{c_N} N_c^{2+\delta})^{\frac{2}{2+\delta}}/N = (\sum_{c=1}^{c_k} (N_c^2/N)^{\frac{2+\delta}{2}} + O(1))^{\frac{2}{2+\delta}} < \infty$ when $N_1^2/N \rightarrow 0$ and c_k is fixed.*

For case (b), using $c_k \leq N$ we have $(\sum_{c=1}^{c_N} N_c^{2+\delta})^{\frac{2}{2+\delta}}/N = (\sum_{c=1}^{c_k} (N_c^2/N)^{\frac{2+\delta}{2}} + o(1))^{\frac{2}{2+\delta}} \leq (N_1^{\frac{2(2+\delta)}{\delta}}/N)^{\frac{2}{2+\delta}} + o(1)$ so our condition (b)(i) implies HL19 rate conditions. Additionally, from the proof of Theorem 2 we can see that under condition (b)(ii) we have $\sum_{c=1}^{c_k} N_c^2/N = O(1)$, thus our conditions imply the ones in HL19 here as well. Thus, in this environment we can view our rate conditions as lower level sufficient conditions for HL19 rate restrictions.

For case (c), when $c_k/c_N \rightarrow 1$, same reasoning as above shows that our rate condition in (iv) implies HL19 conditions. When $c_k/c_N \rightarrow 0$, our rate conditions and HL19 are non-nested. With $\delta = \infty$ (Y has all moments), the condition in (iv) simplifies to $N_1/N \rightarrow 0$, which is equivalent to HL19 general condition $N_1^2/N\lambda_N \rightarrow 0$ in the borderline case $\lambda_N = O(N_1)$. Whereas under $\delta < \infty$ assumption $N_1^{\frac{2+\delta}{\delta}}/N \rightarrow 0$ may imply HL19 rate conditions.

As in Remark 2 we stress that, even when our rate conditions imply HL19 rate conditions, this does not mean that our set of assumptions is nested within theirs as we do not assume uniform integrability.

Remark 5. *Table 1 shows the values of the quantities featured in Theorems 1 and 2 in commonly used network datasets. We can see that the school/village data (first 4 rows) feature mainly sizable ($N_c > 10$), fairly homogenous in size, components, which make them prone to the analysis in Thm. 1 (a), and, indeed N_1/N is small in all of these datasets. Facebook community data (10 groups) correspond more to the case (a) in Thm. 2, but here the condition $N_1^{1+\gamma_1}/N \rightarrow 0$ does not seem to be satisfied even with $\gamma_1 \approx 0$. The community data from other social networks (Youtube, Flickr, LiveJournal) contain significant share (80-90%) of small components, thus correspond to case (b) or (c) of Thm. 2. However, here*

⁷We compare other cases in Theorems 1 and 2 to DMN in Online Supplement B.

Table 1: Network datasets

Dataset	N	Diameter	Avg. degree	c_N/N	c_k/c_N	N_1/N	N_1^2/N
Add Health	75,804	36	8.45	0.001	1.000	0.034	88.29
Karnataka	14,904	7	8.97	0.005	1.000	0.024	8.50
Rice Farmers	7,849	17	5.81	0.006	0.979	0.058	26.49
NJ Schools	27,435	8	13.40	0.002	1.000	0.031	26.90
Facebook	4,089	11	83.24	0.003	1.000	0.253	261.47
Youtube	1,138,499	71	7.41	0.018	0.141	0.235	63,096.46
Flickr	1,715,255	84	24.02	0.039	0.195	0.294	148,685.82
LiveJournal	4,901,928	93	27.03	0.008	0.103	0.378	699,136.98

Note: c_k is defined as the number of components of size larger than 10. For Youtube, Flickr, LiveJournal data we apply a fast-greedy community detection algorithm to identify components, as raw data features a giant component. Online Supplement F contains a detailed description of the data sources.

also the restrictions on the size of the largest component do not seem to be satisfied. This may suggest that the normal approximation may not work very well in these datasets.

3.1.3 Small worlds

Next, we specialise Theorems 1 and 2 to the small-world setup by providing sufficient conditions for the restrictions on the largest component. Let Δ_N^c denote the diameter of component c and note that $\Delta_N = \max_{c=1,\dots,c_N} \Delta_N^c$.

Assumption SW. (a) $d_{max} \geq 2, d_{max} = O(1)$.

(b) $\Delta_N^c \leq \log_{a_c}(b_c N_c)$ for some constants a_c, b_c such that $\inf_c a_c > 1, \inf_c b_c > 0, \sup_c a_c < \infty, \sup_c b_c < \infty$.

Assumption SW(a) imposes sparsity of the network. The lower bound on the maximal degree has a technical nature. Note that sparsity is often imposed as a condition on average degree, namely $1/N \sum_{i=1}^N d_i = O(1)$, which is implied by our condition. Assumption SW(b) imposes the small-world property, namely that the diameter of the subnetwork is (at most) proportional to $\log N_c$, which, in turn, implies that the diameter of the whole network satisfies $\Delta_N \leq \log_a(bN)$ for some finite constants a and b .

Proposition 1. *Under Assumptions CVAR and SW the conditions in (a), (b)(iv) in Theorem 1 and (a), (b), (c)(iv) in Theorem 2 of the form $N_c^r/N \rightarrow 0$ can be replaced by:*

$$\log_{d_{max}-1} a > r.$$

where $a = \inf_c a_c$.

Combined with Theorems 1 and 2, Proposition 1 gives sufficient conditions for a network comprising of small world components to satisfy the central limit theorem. For example, with all components growing at the same rate with equal γ 's, we need $a > d_{max} - 1$. This means CLT is consistent with having diameter of order $\log N$, but the conditions imposed on $\log_{d_{max}-1} a$ restrict the scaling factor for the network diameter relative to the maximal degree and require that, for given N , the diameter is not too large relative to the maximal degree (in other words, $1/\log(a)$ cannot be too large).

We stress that the condition in Proposition 1 is only sufficient, but not necessary. Consider, for example, the collection of schools in Paluck et al. (2016) as a network. This dataset corresponds well to case (a) in Theorem 1 where we have a growing number of similar components. Actually, in this case we would probably expect the size of all components (i.e. schools) to stay approximately the same as the number of pupils N grows so $N_1/N \rightarrow 0$ is satisfied trivially without the need of restricting the diameter relative to maximal degree.

To the best of our knowledge, there is no general formula linking the parameters of a small world model like the Watts-Strogatz small world (SW) model (Watts & Strogatz (1998)) and the Barabási-Albert (BA) preferential attachment model (Barabási & Albert (1999)) to the constants of proportionality in the diameter so it is difficult to translate our conditions on the constant a to the parameters of these models. However, we have some special cases.

Remark 6. For a random graph $G(N, p)$ with $Np \rightarrow \lambda > 1$, we have $\Delta_N = \log N / \log \lambda$. In such random graph let us truncate the degree at some d_{max} and, on top of that, artificially split the largest component, which is of order N in this case, such that it grows at some slower rate. This way the generated network dynamics will fit our setup in part (b) of Theorem 1 and condition (b)(iv) becomes $\log \lambda > \log(d_{max} - 1)^{\frac{2+\delta}{\delta}} \sup_{c \geq 1} (1 + \gamma_c)$ which gives $\lambda > 9$ when $d_{max} = 4, \gamma_c = 1, \delta = \infty$.

Remark 7. If each network component is a regular random graph with degree d , we have $\Delta_N \approx \log N / \log d$ so we can set $a = d$. By Theorem 1(a) and Proposition 1 the sufficient condition for CLT is $d > d - 1$, which is automatically satisfied.

A natural corollary, providing sufficient conditions for a CLT for multiple small-world networks irrespective of the network structure and values of γ_c , follows:

Corollary 1. Let $\{Y_i\}_{i=1}^{\infty}$ be a sequence of mean zero random variables and define $B_N^2 = \text{Var}(\sqrt{NY})$. Under Assumptions CVAR and SW:

$$\frac{\sqrt{NY}}{B_N} \rightarrow^D N(0, 1)$$

as $N \rightarrow \infty$ (conditionally on network evolution) if either of the following holds:

(a) all c_k growing components satisfy:

$$(i) \frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0 \text{ implies } \frac{N_c(N)}{N_1(N)} \rightarrow 0,$$

$$(ii) \frac{N_c(N)^{\gamma_c}}{N_1(N)^{\gamma_1}} \text{ is weakly decreasing in } N,$$

$$(iii) \text{ there exists } M < \infty \text{ such that } \frac{N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}}}{N_{c_N}(N)^{\gamma_N}} < \left(\frac{N_1(\tilde{N})}{N_1(N)} \right)^{\gamma_1} \text{ for all } (N, \tilde{N}) \text{ such that } c_{\tilde{N}} = c_N + 1 \text{ and } N > M, \tilde{N} > M,$$

$$\text{and } \log_{d_{max}-1} a > \frac{2(2+\delta)}{\delta},$$

(b) all components grow at the same rate with equal γ_c 's and $\log_{d_{max}-1} a > 1$.

4 Means of edge-specific characteristics

In this section we provide limit theorems for means of characteristics of edges between nodes. Applications include means of input-output flows in production networks (see e.g. Acemoglu et al. (2012)) or mean difference in socio-economic status between individuals belonging to the same local community (see e.g. Chetty et al. (2022)).

Let Y_{ij} denote the characteristic of an edge between nodes i and j and assume that there are no flows between separate components, i.e. $Y_{ij} = 0$ if nodes i and j are not connected. With this structure we effectively have c_N components with $N_{c,f} = N_c(N_c - 1)$ outcome pairs Y_{ij} and the analysis resembles the one for node-specific means, but now the effective sample size is $N_f = \sum_{c=1}^{c_N} N_{c,f}$. We can define the edge-specific mean by:

$$\bar{Y}_f = \frac{1}{N_f} \sum_{i=1}^N \sum_{j \neq i} Y_{ij} \mathbb{1}\{l(i, j) < \infty\}.$$

Similarly to node-specific means we will assume that flows in separate network components are statistically independent and modify Assumption CVAR to the present context:

Assumption CVAR'. (a) $l(i, k) = \infty$ implies $Y_{ij} \perp Y_{kl}$ for all (j, l) .

(b) There exist $\{\gamma_c\}_{c=1}^{c_N} : 0 \leq \gamma_c \leq 1, \underline{\sigma}^2 > 0, \bar{\sigma}^2 > 0, \delta > 0$ such that:

$$\underline{\sigma}^2 N_{c,f}^{1+\gamma_c} \leq \text{Var} \left(\sum_{i \in C_c} \sum_{j \in C_c: j \neq i} Y_{ij} \right) \leq \bar{\sigma}^2 N_{c,f}^{1+\gamma_c},$$

$$E \left(\sum_{i \in C_c} \sum_{j \in C_c: j \neq i} Y_{ij} \right)^{2+\delta} \leq K_N^{\frac{2+\delta}{2}} N_{c,f}^{(1+\gamma_c)\frac{2+\delta}{2}}.$$

for all c and N , where $K_N = O(1)$.

We have the following result:

Theorem 3. *Let $\{Y_{ij}\}_{i,j=1}^\infty$ be a sequence of mean zero random variables and define $B_{N,f}^2 = \text{Var}(\sqrt{N_f} \bar{Y}_f)$. Under Assumptions CVAR' and SW:*

$$\frac{\sqrt{N_f} \bar{Y}_f}{B_{N,f}} \rightarrow^D N(0, 1)$$

as $N \rightarrow \infty$ (conditionally on network evolution) if

- (a) condition (a) in Theorem 1 holds with $\log_{d_{max}-1} a > 1$,
- (b) conditions (b)(i)-(iii) in Theorem 1 hold and $\log_{d_{max}-1} a > 2(1 + \gamma_c) \frac{2+\delta}{\delta}, \forall c$,
- (c) condition (a) in Theorem 2 holds with $\log_{d_{max}-1} a > 2(1 + \gamma_c), \forall c$,
- (d) condition (b) in Theorem 2 holds with (i)-(ii) replaced by:
 - (ii)', $\frac{N_c(N)^{1+\gamma_c}}{N_1(N)^{1+\gamma_1}} \rightarrow 0$ implies $\frac{N_c(N)}{N_1(N)} \rightarrow 0$, $\frac{N_c(N)^{\gamma_c}}{N_1(N)^{\gamma_1}}$ is weakly decreasing in N , there exists $M < \infty$ such that $\frac{N_{c_{\tilde{N}}}(\tilde{N})^{\gamma_{\tilde{N}}}}{N_{c_N}(N)^{\gamma_N}} < \left(\frac{N_1(\tilde{N})}{N_1(N)}\right)^{\gamma_1}$ for all $(N, \tilde{N})$ such that $c_{\tilde{N}} = c_N + 1$ and $N > M, \tilde{N} > M$, and $\log_{d_{max}-1} a > 2(1 + \gamma_c), \forall c$,
- (e) condition (c) in Theorem 2 holds with part (iv) replaced by: $\log_{d_{max}-1} a > 2(1 + \gamma_c) \frac{2+\delta}{\delta}, \forall c$, when $\frac{c_k}{c_N} \rightarrow 1$ or $\log_{d_{max}-1} a > \frac{2(2+\delta)}{\delta}$.

Theorem 3 can be used for inference once an estimator of $B_{N,f}$ is available. Note that Theorem 3 strengthens conditions of Theorem 1 and 2 due to the fact that in the current setup each network component contains (up to) $N_c(N_c - 1)$ correlated elements compared to N_c before.

A special case of the result in this section is $Y_{ij} = h(Y_i, Y_j)$, where h is a known, bounded function (e.g. $h(Y_i, Y_j) = |Y_i - Y_j|$ in Chetty et al. (2022)).⁸ Then, $\bar{Y}_f$ becomes a U-statistic and Theorem 3 offers an alternative result to Liu (2025), tailored to our many component network and without imposing any weak dependence conditions within components. Note that a typical U-statistic would be scaled by $N(N-1)$ rather than N_f . In Online Supplement C we show that, under such scaling and additional ψ -mixing conditions within components, one can still obtain asymptotic normality.

⁸It is worth noting that in this case Assumption CVAR' restricts the correlation between Y_i 's. For example, it excludes the case when Y_i 's are perfectly correlated within components as this would mean that Y_{ij} 's are degenerately equal to zero.

5 Variance estimation

In this section we suggest estimators of the variance B_N^2 that can be used for inference together with Theorems 1 and 2. One would expect that analogous estimators would work for edge-specific means (Theorem 3) but, for brevity, we do not provide formal analysis of that case.

Since Y_i 's have zero mean: $B_N^2 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N E(Y_i Y_j) \mathbb{1}\{l(i, j) < \infty\}$, thus a natural estimator arises:

$$\hat{B}_N^2 = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N Y_i Y_j \mathbb{1}\{l(i, j) < \infty\}.$$

Note that this is the cluster covariance estimator (CCE). Bester et al. (2011) provide conditions under which it is valid for spatial networks.

Theorem 4. *Let $\{Y_i\}_{i=1}^\infty$ be a sequence of mean zero random variables, Assumptions CVAR and SW hold and let $E|Y_i|^4$ be bounded for all i . Then:*

$$\text{Var}(\hat{B}_N^2 - B_N^2) \rightarrow 0.$$

as $N \rightarrow \infty$ (conditionally on network evolution) if:

- (a) all components grow at the same rate and $\log_{d_{max}-1} a > 3$,
- (b) all components grow (possibly at different rates) and $\log_{d_{max}-1} a > 4$,
- (c) c_k components grow with N and remaining $c_N - c_k$ components have fixed size, c_k is fixed and $\log_{d_{max}-1} a > 2$,
- (d) c_k components grow with N and remaining $c_N - c_k$ components have fixed size, $c_k \rightarrow \infty$ and $\log_{d_{max}-1} a > 4$

The theorem implies consistency of the proposed estimator and can be seen as an extension of the second part of Theorem 2.1 in DMN or Theorem 3 in HL19 to our setup. Recall that, when all components grow at the same rate with N , Theorem 1 allows the size of the largest connected component to grow at a rate arbitrarily close to N but here the allowed rate is not higher than $N^{1/3}$. Again, this is in line with findings in Kojevnikov & Song (2023) for clustered samples. They show that one requires much stricter conditions for variance estimation than for consistent discrimination of the mean.

Although consistent, this estimator does not work well in practice if there are a few large components, as shown by Cameron et al. (2008). Our Monte Carlo simulations confirm that. Similarly, a related HAC estimator in Kojevnikov et al. (2021) also undercovers when there

is a lot of dependence between Y_i 's (i.e. the “autoregressive” parameter is close to 0.5) in a setup with one component. Thus, following Cameron et al. (2008), as an alternative to the CCE estimator we consider the wild clustered bootstrap with Rademacher weights and find that it performs better in our Monte Carlo simulations. As an alternative one may consider randomisation tests of Canay et al. (2017).

6 Monte Carlo simulations

We conduct a Monte Carlo study to verify our central limit theorems and assess the finite-sample performance of the variance estimators proposed in Section 5. The finite-sample properties of the CCE and the wild cluster bootstrap have been extensively studied in the clustering literature (e.g. Cameron et al. (2008), Bester et al. (2011), MacKinnon & Webb (2017), DMN, Canay et al. (2021)). However, these studies primarily focus on settings with multiple growing clusters, corresponding to cases (a) and (b) in Theorem 1. We introduce designs that incorporate a mixture of growing and fixed-size components, corresponding to cases of Theorem 2.

We consider two network generating algorithms: the Watts-Strogatz small world (SW) model (Watts & Strogatz (1998)) and the Barabási-Albert (BA) preferential attachment model (Barabási & Albert (1999)). The first model generates networks with diameters proportional to $\log N$ whereas the second model produces diameters proportional to $\log N$ or $\log N / \log \log N$ depending on the parameters (Bollobás & Riordan (2004)). For most parameter values the BA model implies that the maximal degree of a node grows with N , thus we further “prune” the graph to make sure that the maximal degree is stable: (1) we start with a node with the highest degree and randomly erase superfluous edges, (2) check if the maximal degree satisfies the imposed bound, (3) if not, we go back to step (1) and repeat the procedure.

In terms of the architecture of the network, we consider both many growing components networks (Theorem 1) and growing & fixed-size components networks (Theorem 2). For the first case we generate $c_N = 5, 10, 20, 25$ components for different sample sizes ($N = 100, 500, 1000, 5000$) as follows: (a) equally-sized components; (b) component sizes equal to $\lceil Ne^{4c/c_N} / \sum_{c=1}^{c_N} e^{4c/c_N} \rceil$ for $c = 1, \dots, c_N - 1$ and all remaining nodes included in component c_N . For Theorem 2, we have: (a) 10 components growing at rates proportional to $N^{\{0.45, 0.4, 0.35, 0.3, 0.25, 0.2, 0.15, 0.1, 0.1, 0.1\}}$, remaining nodes in fixed-size components with size drawn from binomial distribution (max size = 10, mean size = 5); (b) growing and fixed-size components generated as in (a) but 30% of nodes always in fixed-size components; (c) growing and fixed-size components generated as in (a) and the share of fixed-size components decreasing with sample size as $0.4 \rightarrow 0.3 \rightarrow 0.2 \rightarrow 0.1$.

We perform 1000 MC repetitions and use 1000 replications for the bootstrap procedures. Online Supplement E shows examples of networks generated by SW and BA models. Note that pruning of the BA model sometimes produces additional separated components.

6.1 Node-specific means

Let $C(i)$ denote the network component containing node i and $\mathcal{N}_{i,l}$ denote the set of neighbours of node i at distance l . The data is generated from the following process:

$$Y_i = \sum_{l \geq 0} \frac{\rho^l}{|\mathcal{N}_{i,l}|} \sum_{j \in \mathcal{N}_{i,l}} \varepsilon_j$$

where:

$$\varepsilon_i = \sqrt{\rho} X_{C(i)} + \sqrt{1 - \rho} v_i$$

and X 's and v 's are drawn from a standardised Bernoulli distribution (with $p = 0.3$), i.i.d. over components and individuals, respectively. This design makes the outcomes dependent on the network structure and its evolution as well as generating the necessary interdependence of outcomes.⁹ Note that ρ governs both the strength of correlation of individual shocks ε and the strength of dependence of node i 's outcome on other members of his network. We consider both strong ($\rho = 0.9$) and medium ($\rho = 0.5$) correlation. We look at 90% coverage and include results with 95% coverage in Online Supplement D. We consider coverage of confidence intervals built using known variance ("oracle"), CCE estimator $\hat{B}_N^2$ ("estim.") and wild cluster bootstrap ("boot.") introduced in Section 5.

Tables 2-3 contain the simulation result for means of node-specific characteristics. The BA model is parametrised by: m – the number of edges added in each step of building the graph, z_a – appeal of nodes that do not have any connections. The algorithm for building an SW network starts with a circle (or, more generally, lattice) graph and "rewires" some of the connections between neighbouring nodes to some more distant nodes, thus is parametrised by: p – probability of rewiring an edge, k – the number of edges per vertex in the initial circle graph.

When we use the known variance the coverage generally converges to the nominal value of 0.9 but convergence seems to be slow for the networks in Thm. 1 (a) and Thm. 2 (a), especially with strong correlation $\rho = 0.9$. For example, even with $N = 5000$ the coverage in the BA model, Thm. 1 (a), is at 0.933 so quite far from the nominal value.

⁹With $\varepsilon_i = v_i$ the sample average, $\bar{Y}$, becomes a weighted average of N independent shocks, so network dependence does not play a role, whereas when $\rho \rightarrow 1$ the sample mean $\bar{Y}$ becomes a (weighted) average of c_N components-specific shocks X_C .

Table 2: Simulated coverage, node-specific means, 90% level, BA model

N	m	z_a	d_{max}	Δ_N	oracle	$\rho = 0.9$ estim.	boot.	oracle	$\rho = 0.5$ estim.	boot.
Thm. 1 (a)										
100	1	2	10	8	0.942	0.786	0.810	0.911	0.801	0.798
500	1	2	10	13	0.901	0.833	0.861	0.904	0.808	0.861
1000	1	2	10	14	0.888	0.882	0.890	0.905	0.860	0.897
5000	1	2	10	19	0.914	0.872	0.903	0.893	0.850	0.875
100	2	2	12	5	0.968	0.808	0.806	0.923	0.776	0.794
500	2	2	19	5	0.926	0.798	0.827	0.930	0.813	0.848
1000	2	2	19	6	0.914	0.841	0.887	0.904	0.855	0.889
5000	2	2	20	7	0.933	0.860	0.851	0.925	0.855	0.901
Thm. 1 (b)										
100	1	2	10	9	0.902	0.550	0.800	0.914	0.549	0.838
500	1	2	10	13	0.930	0.671	0.826	0.931	0.675	0.840
1000	1	2	10	14	0.915	0.753	0.867	0.904	0.797	0.861
5000	1	2	10	18	0.901	0.852	0.883	0.898	0.854	0.892
100	2	2	17	4	0.890	0.493	0.770	0.904	0.467	0.779
500	2	2	20	6	0.924	0.605	0.844	0.919	0.687	0.838
1000	2	2	20	6	0.926	0.781	0.869	0.911	0.798	0.858
5000	2	2	20	9	0.896	0.779	0.862	0.905	0.813	0.896
Thm. 2 (a)										
100	1	2	8	8	0.925	0.621	0.827	0.920	0.696	0.854
500	1	2	10	13	0.926	0.655	0.826	0.923	0.696	0.845
1000	1	2	10	15	0.925	0.716	0.855	0.920	0.754	0.881
5000	1	2	10	18	0.918	0.824	0.863	0.928	0.856	0.861
100	2	2	10	5	0.934	0.687	0.853	0.927	0.687	0.863
500	2	2	20	6	0.930	0.668	0.813	0.922	0.683	0.827
1000	2	2	20	7	0.922	0.706	0.818	0.921	0.742	0.849
5000	2	2	20	7	0.926	0.802	0.837	0.926	0.793	0.865
Thm. 2 (b)										
100	1	2	7	9	0.904	0.571	0.866	0.927	0.663	0.878
500	1	2	10	14	0.933	0.670	0.852	0.907	0.750	0.867
1000	1	2	10	15	0.891	0.800	0.863	0.905	0.815	0.889
5000	1	2	10	20	0.890	0.850	0.868	0.911	0.843	0.897
100	2	2	15	5	0.914	0.596	0.872	0.924	0.653	0.873
500	2	2	20	6	0.928	0.684	0.824	0.928	0.719	0.869
1000	2	2	20	6	0.916	0.776	0.873	0.921	0.783	0.859
5000	2	2	20	9	0.913	0.804	0.861	0.903	0.819	0.879
Thm. 2 (c)										
100	1	2	6	10	0.919	0.615	0.868	0.919	0.693	0.882
500	1	2	10	14	0.918	0.652	0.884	0.916	0.698	0.885
1000	1	2	10	18	0.904	0.765	0.882	0.923	0.787	0.885
5000	1	2	10	20	0.885	0.846	0.860	0.898	0.851	0.913
100	2	2	11	5	0.914	0.650	0.853	0.910	0.686	0.855
500	2	2	18	6	0.931	0.680	0.836	0.917	0.702	0.845
1000	2	2	20	6	0.906	0.788	0.876	0.895	0.796	0.862
5000	2	2	20	9	0.910	0.814	0.861	0.901	0.808	0.880

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. “Oracle” – known variance, “estim.” – variance estimator $\hat{B}_N^2$, “boot.” – wild cluster bootstrap.

Note that networks in Thm. 1 (a) feature a small number of large components so with strong correlation and shocks drawn from 0/1 there is little variation in the data even with $N = 5000$ (25 components), hence very large sample sizes will be required for attaining the CLT-implied coverage. In the design for Thm. 2 (a) we have relatively large number of components ($c_N \approx 800$ for $N = 5000$) so the poor coverage is mainly caused by limited variation in the data with relatively strong correlation within components and Bernoulli shocks. This is confirmed by MC results with binomial shocks in Online Supplement D, a

Table 3: Simulated coverage, node-specific means, 90% level, SW model

N	p	k	d_{max}	Δ_N	oracle	$\rho = 0.9$ estim.	boot.	oracle	$\rho = 0.5$ estim.	boot.
Thm. 1 (a)										
100	0.05	2	5	5	0.966	0.773	0.826	0.915	0.796	0.795
500	0.05	2	7	11	0.938	0.810	0.831	0.926	0.802	0.847
1000	0.05	2	7	10	0.914	0.851	0.890	0.921	0.846	0.867
5000	0.05	2	7	16	0.908	0.863	0.887	0.911	0.870	0.892
100	0.1	10	19	1	0.964	0.763	0.777	0.919	0.772	0.814
500	0.1	10	25	3	0.926	0.811	0.836	0.936	0.811	0.833
1000	0.1	10	28	3	0.918	0.846	0.884	0.913	0.856	0.876
5000	0.1	10	28	4	0.923	0.880	0.861	0.905	0.869	0.883
Thm. 1 (b)										
100	0.05	2	5	9	0.914	0.429	0.816	0.926	0.509	0.805
500	0.05	2	7	12	0.939	0.648	0.834	0.938	0.647	0.826
1000	0.05	2	7	13	0.905	0.760	0.836	0.920	0.749	0.860
5000	0.05	2	7	18	0.911	0.810	0.878	0.892	0.821	0.859
100	0.1	10	24	3	0.913	0.477	0.810	0.911	0.488	0.829
500	0.1	10	26	3	0.914	0.617	0.829	0.924	0.649	0.832
1000	0.1	10	29	3	0.899	0.739	0.857	0.898	0.774	0.864
5000	0.1	10	26	4	0.891	0.803	0.889	0.916	0.828	0.871
Thm. 2 (a)										
100	0.05	2	5	5	0.924	0.649	0.842	0.913	0.691	0.845
500	0.05	2	7	13	0.940	0.702	0.829	0.928	0.682	0.825
1000	0.05	2	7	10	0.939	0.686	0.849	0.923	0.742	0.841
5000	0.05	2	7	16	0.941	0.754	0.848	0.926	0.773	0.837
100	0.1	10	23	2	0.932	0.691	0.862	0.915	0.688	0.860
500	0.1	10	25	3	0.924	0.670	0.839	0.946	0.687	0.812
1000	0.1	10	25	4	0.939	0.697	0.840	0.932	0.740	0.850
5000	0.1	10	27	4	0.921	0.777	0.836	0.930	0.808	0.867
Thm. 2 (b)										
100	0.05	2	6	8	0.923	0.518	0.857	0.919	0.612	0.881
500	0.05	2	6	12	0.915	0.635	0.823	0.929	0.717	0.847
1000	0.05	2	7	11	0.901	0.781	0.848	0.913	0.816	0.877
5000	0.05	2	7	17	0.913	0.802	0.862	0.903	0.828	0.853
100	0.1	10	24	2	0.920	0.608	0.849	0.911	0.650	0.864
500	0.1	10	26	3	0.939	0.664	0.823	0.924	0.716	0.859
1000	0.1	10	26	3	0.919	0.793	0.852	0.924	0.810	0.891
5000	0.1	10	27	4	0.899	0.811	0.867	0.900	0.827	0.884
Thm. 2 (c)										
100	0.05	2	6	7	0.913	0.527	0.859	0.912	0.643	0.869
500	0.05	2	8	12	0.930	0.677	0.829	0.939	0.759	0.855
1000	0.05	2	8	15	0.911	0.792	0.851	0.895	0.791	0.874
5000	0.05	2	7	17	0.909	0.816	0.870	0.914	0.825	0.874
100	0.1	10	25	2	0.921	0.607	0.872	0.912	0.653	0.862
500	0.1	10	25	3	0.927	0.682	0.822	0.908	0.709	0.822
1000	0.1	10	26	3	0.912	0.741	0.843	0.911	0.774	0.868
5000	0.1	10	28	4	0.907	0.802	0.835	0.896	0.847	0.867

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. “Oracle” – known variance, “estim.” – variance estimator $\hat{B}_N^2$, “boot.” – wild cluster bootstrap.

design implying a larger variation in individual outcomes, which features empirical coverage that is much closer to the nominal values for Thm. 2 (a) (and Thm. 1 (a)). Additionally, the 95% coverage tables therein show better coverage. Overall, the results imply that with network dependence one may need samples of $N = 5000$ or even beyond for CLT to provide an accurate approximation.

Once we use the estimated variance $\hat{B}_N^2$ (CCE) the coverage deteriorates substantially with values somehow close to the nominal values only for Thm. 1 (a) where we have ho-

mogenous components. This is in line with other MC studies in the literature. As expected, the wild cluster bootstrap improves on CCE, obtaining coverage close to nominal for most of the designs in Thm. 1 (a) with $N = 5000$. However, it noticeably undercovers in all other cases, where we have highly heterogeneous cluster sizes, e.g. 0.836 coverage for Thm. 2 (a) and 0.835 for Thm. 2 (c) even with $N = 5000$. This is in line with the MC simulation results in the clustering literature (DMN, MacKinnon et al. (2023)), which stress heterogeneity of cluster sizes as a factor affecting performance of the cluster bootstrap.

Overall, the coverage rates improve when correlation between outcomes decreases ($\rho = 0.9$ versus $\rho = 0.5$), which confirms that larger sample sizes will be needed to justify CLT approximation with stronger intra-components dependence.

6.2 Edge-specific means

We define the edge-specific outcomes as $Y_{ij} = |Y_i - Y_j|$, where Y_i 's are generated as before. This leads to non-zero correlation between any Y_{ij} and Y_{kl} belonging to the same component. As variance estimation in the U-statistic setup is more involved than with simple means and we do not provide variance estimators above, we only provide coverage values with known variance.¹⁰ Note that here increasing ρ may reduce the variance of Y_{ij} . In an extreme example of $\rho = 1$ and a regular network (e.g. "circle") the edge-specific outcomes are all equal to zero, which violates assumption CVAR'.

Table 4 shows that for all specifications of the network evolution the coverage probabilities are generally getting closer to 90% when N increases, in line with our CLT in Thm. 3. Interestingly, with strong correlation ($\rho = 0.9$) the coverage is generally better for designs in Thm. 1 than Thm 2. Intuitively, this is expected. Designs in Thm. 2 generally feature large number of small components but Y_i 's within these components are highly correlated so Y_{ij} 's are close to zero and they do not contribute to variation in $\bar{Y}^f$, which is mainly determined by larger subnetworks. However, the larger subnetworks in Thm. 2 are smaller than the ones in Thm. 1 and, thus, also Y_{ij} 's are less variable there – note that Y_i is affected by more v_j shocks, the larger the component is. Overall, this means less variation in edge-specific outcomes in designs in Thm. 2 and poorer coverage there. This mechanism is, however, weakened with lower ρ so we see similar coverage for Thm. 1 and Thm. 2 with $\rho = 0.5$.

7 Conclusion

Many social and economic networks are sparse and are small-world. We show that data coming from (multiple) such networks satisfies a central limit theorem under the additional assumption restricting the constant of proportionality of the diameter to $\log N$.

¹⁰We approximate the variance using 5000 simulations.

Table 4: Simulated coverage, edge-specific means, 90% level

N	BA model						SW model					
	m	z_a	d_{max}	Δ_N	$\rho = 0.9$ oracle	$\rho = 0.5$ oracle	p	k	d_{max}	Δ_N	$\rho = 0.9$ oracle	$\rho = 0.5$ oracle
Thm. 1 (a)												
100	1	2	10	8	0.889	0.904	0.05	2	5	5	0.913	0.897
500	1	2	10	13	0.917	0.890	0.05	2	7	11	0.892	0.885
1000	1	2	10	14	0.916	0.914	0.05	2	7	10	0.891	0.919
5000	1	2	10	19	0.898	0.917	0.05	2	7	16	0.895	0.889
100	2	2	12	5	0.917	0.901	0.1	10	19	1	0.916	0.917
500	2	2	19	5	0.910	0.874	0.1	10	25	3	0.881	0.901
1000	2	2	19	6	0.906	0.886	0.1	10	28	3	0.909	0.908
5000	2	2	20	7	0.912	0.891	0.1	10	28	4	0.889	0.899
Thm. 1 (b)												
100	1	2	10	9	0.923	0.885	0.05	2	5	9	0.904	0.913
500	1	2	10	13	0.909	0.907	0.05	2	7	12	0.930	0.902
1000	1	2	10	14	0.917	0.906	0.05	2	7	13	0.913	0.900
5000	1	2	10	18	0.899	0.902	0.05	2	7	18	0.904	0.888
100	2	2	17	4	0.907	0.908	0.1	10	24	3	0.909	0.922
500	2	2	20	6	0.915	0.885	0.1	10	26	3	0.912	0.910
1000	2	2	20	6	0.920	0.902	0.1	10	29	3	0.897	0.897
5000	2	2	20	9	0.915	0.881	0.1	10	26	4	0.908	0.889
Thm. 2 (a)												
100	1	2	8	8	0.906	0.903	0.05	2	5	5	0.922	0.899
500	1	2	10	13	0.914	0.910	0.05	2	7	13	0.919	0.881
1000	1	2	10	15	0.930	0.899	0.05	2	7	10	0.917	0.893
5000	1	2	10	18	0.901	0.901	0.05	2	7	16	0.927	0.904
100	2	2	10	5	0.940	0.906	0.1	10	23	2	0.908	0.908
500	2	2	20	6	0.912	0.908	0.1	10	25	3	0.908	0.877
1000	2	2	20	7	0.922	0.894	0.1	10	25	4	0.891	0.896
5000	2	2	20	7	0.933	0.913	0.1	10	27	4	0.896	0.912
Thm. 2 (b)												
100	1	2	7	9	0.911	0.900	0.05	2	6	8	0.899	0.892
500	1	2	10	14	0.927	0.891	0.05	2	6	12	0.909	0.899
1000	1	2	10	15	0.921	0.889	0.05	2	7	11	0.911	0.901
5000	1	2	10	20	0.936	0.901	0.05	2	7	17	0.900	0.910
100	2	2	15	5	0.910	0.911	0.1	10	24	2	0.915	0.912
500	2	2	20	6	0.925	0.915	0.1	10	26	3	0.890	0.912
1000	2	2	20	6	0.897	0.905	0.1	10	26	3	0.906	0.906
5000	2	2	20	9	0.913	0.897	0.1	10	27	4	0.920	0.890
Thm. 2 (c)												
100	1	2	6	10	0.924	0.889	0.05	2	6	7	0.921	0.913
500	1	2	10	14	0.922	0.883	0.05	2	8	12	0.927	0.893
1000	1	2	10	18	0.942	0.891	0.05	2	8	15	0.920	0.895
5000	1	2	10	20	0.910	0.899	0.05	2	7	17	0.915	0.907
100	2	2	11	5	0.902	0.919	0.1	10	25	2	0.878	0.921
500	2	2	18	6	0.911	0.901	0.1	10	25	3	0.915	0.896
1000	2	2	20	6	0.904	0.905	0.1	10	26	3	0.897	0.892
5000	2	2	20	9	0.932	0.909	0.1	10	28	4	0.908	0.904

Note: 1000 Monte Carlo simulations, 1000 bootstrap replications. “Oracle” – known variance.

We consider a simple setup of undirected unweighted networks but the results should extend naturally to directed networks and networks in which we can assign (bounded) weights to covariances of characteristics between two connected nodes. If these weights would vanish or decrease sufficiently fast between large connected components, then one could potentially be able to extend our results to networks with a giant component of size $O(N)$, i.e. larger than allowed in our current setup.

Similarly, a natural extension would be to allow all network components to be "weakly" connected, for example by having low conductance (see Leung (2023)) and weak depen-

dence between the clusters. If a CLT would obtain in such setting with unrestricted intra-component dependence remains an open question.

References

- Acemoglu, D., Carvalho, V. M., Ozdaglar, A. & Tahbaz-Salehi, A. (2012), ‘The network origins of aggregate fluctuations’, *Econometrica* **80**(5), 1977–2016.
- Baldi, P. & Rinott, Y. (1989), ‘On normal approximations of distributions in terms of dependency graphs’, *The Annals of Probability* **17**(4), 1646 – 1650.
- Barabási, A.-L. & Albert, R. (1999), ‘Emergence of scaling in random networks’, *Science* **286**(5439), 509–512.
- Bester, C. A., Conley, T. G. & Hansen, C. B. (2011), ‘Inference with dependent data using cluster covariance estimators’, *Journal of Econometrics* **165**(2), 137–151.
- Bickel, P., Choi, D., Chang, X. & Zhang, H. (2013), ‘Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels’, *The Annals of Statistics* **41**(4), 1922–1943.
- Bickel, P. J., Chen, A. & Levina, E. (2011), ‘The method of moments and degree distributions for network models’, *The Annals of Statistics* **39**(5), 2280 – 2301.
- Bollobás, B. & Riordan, O. (2004), ‘The diameter of a scale-free random graph’, *Combinatorica* **24**(1), 5–34.
- Cameron, A. C., Gelbach, J. B. & Miller, D. L. (2008), ‘Bootstrap-based improvements for inference with clustered errors’, *The Review of Economics and Statistics* **90**(3), 414–427.
- Canay, I. A., Romano, J. P. & Shaikh, A. M. (2017), ‘Randomization tests under an approximate symmetry assumption’, *Econometrica* **85**(3), 1013–1030.
- Canay, I. A., Santos, A. & Shaikh, A. M. (2021), ‘The wild bootstrap with a “small” number of “large” clusters’, *The Review of Economics and Statistics* **103**(2), 346–363.
- Chen, L. H. Y. & Shao, Q.-M. (2004), ‘Normal approximation under local dependence’, *The Annals of Probability* **32**(3), 1985–2028.
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebe, J., Hendren, N., Fluegge, R. B., Gong, S., Gonzalez, F., Grondin, A., Jacob, M., Johnston, D., Koenen, M., Laguna-Muggenburg, E., Mudekereza, F., Rutter, T., Thor, N., Townsend, W., Zhang, R., Bailey, M., Barberá, P., Bhole, M. & Wernerfelt, N. (2022), ‘Social capital I: measurement and associations with economic mobility’, *Nature* **608**(7921), 108–121.
- Djogbenou, A. A., MacKinnon, J. G. & Nielsen, M. Ø. (2019), ‘Asymptotic theory and wild bootstrap inference with clustered errors’, *Journal of Econometrics* **212**(2), 393–412.
- Hall, P. & Heyde, C. (1980), *Martingale Limit Theory and its Application*, Probability and Mathematical Statistics: A Series of Monographs and Textbooks, Academic Press.

- Hansen, B. E. & Lee, S. (2019), ‘Asymptotic theory for clustered samples’, *Journal of Econometrics* **210**(2), 268–290.
- Harris, K. M., Halpern, C. T., Smolen, A. & Haberstick, B. C. (2006), ‘The national longitudinal study of adolescent health (Add Health) twin data’, *Twin Research and Human Genetics* **9**(6), 988–997.
- Jackson, M. O. (2008), *Social and economic networks*, Princeton University Press, Princeton, N.J. ; Woodstock.
- Jenish, N. & Prucha, I. R. (2012), ‘On spatial processes and asymptotic inference under near-epoch dependence’, *Journal of Econometrics* **170**(1), 178–190.
- Kojevnikov, D., Marmer, V. & Song, K. (2021), ‘Limit theorems for network dependent random variables’, *Journal of Econometrics* **222**(2), 882–908.
- Kojevnikov, D. & Song, K. (2023), ‘Some impossibility results for inference with cluster dependence with large clusters’, *Journal of Econometrics* **237**(2, Part A), 105524.
- Kolaczyk, E. (2009), *Statistical Analysis of Network Data: Methods and Models*, Springer.
- Kuersteiner, G. M. (2019), Limit theorems for data with network structure. Working paper.
- Kuersteiner, G. M. & Prucha, I. R. (2013), ‘Limit theory for panel data models with cross sectional dependence and sequential exogeneity’, *Journal of Econometrics* **174**(2), 107–126.
- Leung, M. P. (2023), ‘Network cluster-robust inference’, *Econometrica* **91**(2), 641–667.
- Leung, M. P. & Moon, H. R. (2026), ‘Normal approximation in large network models’, *The Review of Economic Studies* .
- Liu, W. (2025), ‘Normal approximation for U-statistics with cross-sectional dependence’. Working paper.
- MacKinnon, J. G., Nielsen, M. Ø. & Webb, M. D. (2023), ‘Cluster-robust inference: A guide to empirical practice’, *Journal of Econometrics* **232**(2), 272–299.
- MacKinnon, J. G. & Webb, M. D. (2017), ‘Wild bootstrap inference for wildly different cluster sizes’, *Journal of Applied Econometrics* **32**(2), pp. 233–254.
- Matsushita, Y. & Otsu, T. (2024), ‘Empirical likelihood for network data’, *Journal of the American Statistical Association* **119**(547), 2117–2128.
- Ogburn, E. L., Sofrygin, O., Díaz, I. & van der Laan, M. J. (2024), ‘Causal inference for social network data’, *Journal of the American Statistical Association* **119**(545), 597–611.
- Paluck, E. L., Shepherd, H. & Aronow, P. M. (2016), ‘Changing climates of conflict: A social network experiment in 56 schools’, *Proceedings of the National Academy of Sciences* **113**(3), 566–571.

- Pineda-Villavicencio, G. & Wood, D. R. (2015), ‘The degree-diameter problem for sparse graph classes’, *The Electronic Journal of Combinatorics* **22**(2), 1–20.
- Romano, J. P. & Wolf, M. (2000), ‘A more general central limit theorem for m-dependent random variables with unbounded m’, *Statistics & Probability Letters* **47**(2), 115–124.
- Tomassini, M. & Luthi, L. (2007), ‘Empirical analysis of the evolution of a scientific collaboration network’, *Physica A: Statistical Mechanics and its Applications* **385**(2), 750–764.
- Ugander, J., Karrer, B., Backstrom, L. & Marlow, C. (2011), The anatomy of the Facebook social graph. Working Paper.
- Viennet, G. (1997), ‘Inequalities for absolutely regular sequences: application to density estimation’, *Probability Theory and Related Fields* **107**(4), 467–492.
- Watts, D. J. (1999), *Small worlds: the dynamics of networks between order and randomness*, Princeton studies in complexity, Princeton University Press, Princeton, N.J.
- Watts, D. J. & Strogatz, S. H. (1998), ‘Collective dynamics of “small-world” networks’, *Nature* **393**(6684), 440–442.