NETWORK DIFFERENTIAL CONNECTIVITY ANALYSIS

BY SEN ZHAO∗, AND ALI SHOJAIE†,

Google Research∗ and University of Washington†

Identifying differences in networks has become a canonical problem in many biological applications. Existing methods try to accomplish this goal by either directly comparing the estimated structures of two networks, or testing the null hypothesis that the covariance or inverse covariance matrices in two populations are identical. However, estimation approaches do not provide measures of uncertainty, e.g., p-values, whereas existing testing approaches could lead to misleading results, as we illustrate in this paper. To address these shortcomings, we propose a qualitative hypothesis testing framework, which tests whether the connectivity structures in the two networks are the same. Our framework is especially appropriate if the goal is to identify nodes or edges that are differentially connected. No existing approach could test such hypotheses and provide corresponding measures of uncertainty. Theoretically, we show that under appropriate conditions, our proposal correctly controls the type-I error rate in testing the qualitative hypothesis. Empirically, we demonstrate the performance of our proposal using simulation studies and applications in cancer genomics.

1. Introduction. Changes in biological networks, such as gene regulatory and brain connectivity networks, have been found to associate with the onset and progression of complex diseases (see, e.g., Bassett and Bullmore, 2009; Barabási, Gulbahce and Loscalzo, 2011). Locating differentially connected nodes and edges in the network of diseased and healthy individuals—referred to as differential network biology (Ideker and Krogan, 2012)—can help researchers delineate underlying disease mechanism; network-based biomarkers can also serve as effective diagnostic tools and guide new therapies.

Consider the networks of diseased and healthy individuals, \( G^I = (\mathcal{V}, \mathcal{E}^I) \) and \( G^{II} = (\mathcal{V}, \mathcal{E}^{II}) \), defined on the common set of nodes \( \mathcal{V} = \{1, \ldots, p\} \). Let \( ne^m_j \) be the neighborhood of node \( j \) in network \( G^m \), i.e.,

\[
ne^m_j \equiv \{k \neq j : (j, k) \in \mathcal{E}^m\}, \quad m \in \{I, II\}.
\]

The scientists’ quest to identify differences in the two networks corresponds
to testing

\begin{equation}
H_{0,j}^* : ne_j^I = ne_j^II \quad \text{versus} \quad H_{a,j}^* : ne_j^I \neq ne_j^II.
\end{equation}

Under $H_{0,j}^*$, node $j$ is connected to the same set of nodes in both networks. In biological applications, the edge sets $\mathcal{E}^I$ and $\mathcal{E}^II$ are often not directly observed; rather, they are estimated from observations $X^I$ and $X^II$, often using graphical modeling tools (see, e.g., Krumsieck et al., 2011; Meinshausen and Bühlmann, 2006; Friedman, Hastie and Tibshirani, 2008). Under the assumption that $X^I$ and $X^II$ both follow $p$-variate normal distributions, we can infer the conditional independence relationships among variables from entries of inverse covariance matrices. Formally, nodes $j,k \in V$ are connected in network $G^m$ if and only if $\Omega_{jk}^m \neq 0$, where $\Omega^m$ denotes the inverse covariance matrix of variables in population $m$. One drawback of these methods is that they estimate networks independently in each population, whereas biological networks in related subpopulations are expected to share many edges (see, e.g., Kitano, 2004). Motivated by this similarity, recent approaches for joint estimation of multiple graphical models (e.g., Guo et al., 2011; Danaher, Wang and Witten, 2014; Zhao, Cai and Li, 2014; Peterson, Stingo and Vannucci, 2015; Saegusa and Shojaie, 2016) encourage the estimated networks to have common edges, while allowing for differences between the networks.

Using gene expression data from the Cancer Genome Atlas (TCGA), we applied the joint graphical lasso (JGL) procedure of Danaher, Wang and Witten (2014) to estimate the genetic networks of ER negative (ER-) and ER positive (ER+) subtypes of breast cancer from $n^I = 117$ and $n^II = 407$ ER- and ER+ samples, respectively. We focused on $p = 358$ genes in cancer related pathways from KEGG. The estimated differentially connected network of ER- and ER+ samples is shown in Figure 1. For this figure, the tuning parameters for the lasso and group lasso penalties were set to 0.5 and 0.05, respectively. As expected, the two estimated networks share many common edges, but also have noticeable differences. Unfortunately, while these results are insightful for exploratory analyses, they do not provide measures of uncertainty, e.g., $p$-values, and are thus of limited utility for drawing scientific conclusions.

Measures of uncertainty for a single inverse covariance matrix are provided by a few recent hypothesis testing procedures (Ren et al., 2015; Janková and van de Geer, 2015, 2017; Xia and Li, 2017). These methods examine the null hypothesis $\Omega_{jk}^m = 0, m \in \{I, II\}$ for $j \neq k$, and hence could control the probability of falsely detecting a nonexistent edge in both networks. However, they do not control the false positive rates of $H_{0,j}^*$. This is because
Fig 1. Differentially connected edges between ER- and ER+ breast cancer patients found by group graphical lasso. Yellow edges are genetic interactions that are found in ER- but not ER+ breast cancer patients; gray edges are genetic interactions that are found in ER+ but not ER- breast cancer patients. Identically connected edges are omitted.

$H_{0,j}$ concerns the coexistence of edges in the two networks. Thus, the false positive rate of $H_{0,j}$ not only depends on the probability of falsely detecting a nonexistent edge, but also depends on the probability of correctly detecting an existent edge. While single network hypothesis testing methods control the former probability, they do not control the latter.

To address the above-mentioned issue and identify differences between two high-dimensional GGMs, a number of recent inference procedures examines whether the corresponding entries in two inverse covariance matrices are equal. For example, Xia, Cai and Cai (2015) tests whether $\Omega_{I,jk} = \Omega_{II,jk}$, Belilovsky, Varoquaux and Blaschko (2016) tests whether $\Omega_{I,jk}/\Omega_{I,jj} = \Omega_{II,jk}/\Omega_{II,jj}$, and Städler and Mukherjee (2016) tests whether $\phi^I = \phi^{II}$, where $\phi$ parametrizes the underlying data generation distribution. A permutation-based approach was also proposed by Gill, Datta and Datta (2014).

We applied the method of Gill, Datta and Datta (2014), implemented in the dna R-package, to the TCGA gene expression data described above. Controlling the family-wise error rate at 0.1 level using Holm’s procedure, dna with 100 permutations found 75 genes (out of 358 genes) to be differ-
entially connected in ER+ and ER- groups. Unfortunately, dna is unable to infer the direction of differential connectivity, i.e., whether a node has more connections in ER- or ER+ samples; it is also unable to infer differentially connected edges. Furthermore, dna seems to be highly sensitive to the number of permutations: with 10 permutations, 170 genes are found to be differentially connected at 0.1 family-wise error rate level, whereas with 1000 permutation samples, only 28 genes are found to be differentially connected.

In addition to the above limitations of dna, in Section 2, we argue that quantitative testing procedures that focus on differences in values of the inverse covariance matrices may more generally lead to misleading conclusions about structural differences in two networks. To circumvent this problem, in this paper we propose a testing framework that directly examines whether the two networks have the same connectivity patterns. This framework for high-dimensional data, termed differential connectivity analysis (DCA), identifies differentially connected nodes or edges in two networks by testing the qualitative hypothesis that supports of two inverse covariance matrices are the same. More specifically, we recast $H^*_{0,j}$ and $H^*_{a,j}$ as the following equivalent hypotheses

\begin{align}
H^*_{0,j} & : \text{supp} \left( \Omega_{I,j}^m \right) = \text{supp} \left( \Omega_{II,j}^m \right), \tag{3} \\
H^*_{a,j} & : \text{supp} \left( \Omega_{I,j}^m \right) \neq \text{supp} \left( \Omega_{II,j}^m \right), \tag{4}
\end{align}

where $\Omega_{m,j}^m$ denotes the $j$th column of the inverse covariance matrix in population $m$. As a comparison to the quantitative inference procedure of Gill, Datta and Datta (2014), our DCA procedure identified 60 differentially connected genes in ER- and ER+ networks (see Figure 3). Among the differentially connected genes identified by DCA, 48 (80%) are also identified by JGL, whereas 47 of the 75 (63%) differentially connected genes identified by dna overlap with JGL findings. While differential edges identified by JGL have uncharacterized uncertainty, the aggregate results at the node levels are less random. Thus, the larger agreement of DCA and JGL results may suggest more informative findings. We will revisit and expand this data analysis in Section 4.

After discussing the challenges of differential network analysis using quantitative inference procedures in Section 2, we present our new qualitative inference framework, DCA, in Section 3, where we also discuss specific procedures using lasso neighborhood selection. To the best of our knowledge, DCA is the first inference framework that can formally test structural differences in two networks. Moreover, DCA is a general framework that can incorporate various estimation and hypothesis testing methods. Finally, while in this paper we present the results for the case of Gaussian data, our re-
RESULTS are valid more generally for sub-Gaussian data; the only difference is that without Gaussianity, the network encodes conditional correlation, rather than conditional dependence.

2. Challenges of Differential Network Analysis. In this section, we discuss the challenges of testing $H^*_{0,j}: ne^I_j = ne^II_j$ and why existing quantitative inference procedures may not be able to test this hypothesis.

As mentioned in the previous section, recent proposals for differential network analysis (Gill, Datta and Datta, 2014; Xia, Cai and Cai, 2015; Belilovsky, Varoquaux and Blaschko, 2016; Städler and Mukherjee, 2016) examine whether corresponding entries in two inverse covariance matrices are equal. While this seems natural, the primary limitation of these methods is that examining differences in magnitudes of $\Omega^I$ and $\Omega^II$ may lead to misleading conclusions. Consider the following illustrative example with three Gaussian variables: suppose in population I, variable 1 causally affects variables 2 and 3, and variable 2 causally affects variable 3. Suppose, in addition, that in population II, the effect of variable 1 on variable 2 remains intact, while the effect of variables 1 and 2 on variable 3 disappears. The undirected networks corresponding to the two GGMs are portrayed in Figure 2.

Suppose, without loss of generality, that the inverse covariance matrix of variables in population I is

$$\Omega^I = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix}.$$ 

Further, suppose that $x^II_1$ has the same distribution as $x^I_1$, i.e., $x^II_1 \sim_d x^I_1$. The unchanged (causal) relationship of variables 1 and 2 leads to $x^II_2 \sim_d x^I_2$. On the other hand, $x^II_3$ is independent of $x^II_1$ and $x^II_2$ due to its disappeared

---

**Fig 2.** Conditional dependency structures of variables in populations I and II.
relationship with variables 1 and 2. We can verify that
\[
\Omega^{II} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.75 & 0 \\ 0 & 0 & \text{Var}(x_3^{II}) \end{bmatrix}.
\]

This is because by the normality of \(X^I\) and the fact that
\[
\Omega^I = \begin{bmatrix} 1 & 0.5 & 0.5 \\ 0.5 & 1 & 0.5 \\ 0.5 & 0.5 & 1 \end{bmatrix},
\]
we have \(X^I \sim_{i.i.d.} \mathcal{N}_3(0, \Sigma^I)\), where
\[
\Sigma^I = (\Omega^I)^{-1} = \begin{bmatrix} 1.5 & -0.5 & -0.5 \\ -0.5 & 1.5 & -0.5 \\ -0.5 & -0.5 & 1.5 \end{bmatrix}.
\]

In population II, we have \(x_1^{II} = d x_1^I, x_2^{II} = d x_2^I\) and \(x_3^{II} \perp \perp x_{1,2}^{II}\). Thus, \(X^{II} \sim_{i.i.d.} \mathcal{N}_3(0, \Sigma^{II})\), where
\[
\Sigma^{II} = \begin{bmatrix} 1.5 & -0.5 & 0 \\ -0.5 & 1.5 & 0 \\ 0 & 0 & \text{Var}(x_3^{II}) \end{bmatrix},
\]
which implies that
\[
\Omega^{II} = (\Sigma^{II})^{-1} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.75 & 0 \\ 0 & 0 & 1/\text{Var}(x_3^{II}) \end{bmatrix}.
\]

In this example, although the relationship between \(x_1\) and \(x_2\) is the same in the two networks, \(\Omega_{1,2}^I \neq \Omega_{1,2}^{II}\) and \(\Omega_{1,2}^I/\Omega_{1,1}^I \neq \Omega_{1,2}^{II}/\Omega_{1,1}^{II}\). Thus, existing quantitative tests (e.g., Gill, Datta and Datta, 2014; Xia, Cai and Cai, 2015; Belilovsky, Varoquaux and Blaschko, 2016; Städler and Mukherjee, 2016) would falsely detect \((1, 2)\) as a differentially connected edge!

To further illustrate that quantitative tests may fail to control the type-I error rate of qualitative hypotheses \(H_{0,j}^* : ne_j^I = ne_j^{II}\), we examine how the permutation test of Gill, Datta and Datta (2014) controls the type-I error rate for \(H_{0,j}^*\) in the simulation setting of Section 5. The type-I error rates for node-wide tests of differential connectivity are shown in Table 1.

The errors in Table 1 seem unbelievably large. But note that even on qualitatively identically connected nodes, the connection strength as reflected by
inverse covariance may be vastly different on many edges due to the difference in qualitative connectivity of other nodes, a phenomenon illustrated in the above example. As a result, any quantitative test with sufficient power will falsely conclude that those qualitatively identically connected nodes are differentially connected. That is exactly the issue highlighted in the above illustrative example: Tests for equality of inverse covariance values (e.g., Gill, Datta and Datta, 2014; Xia, Cai and Cai, 2015; Belilovsky, Varoquaux and Blaschko, 2016; Städler and Mukherjee, 2016) were designed to identify quantitative differences, and should not be used to identify qualitative differences in two networks, or differential connectivity, if that is indeed the scientific question.


3.1. Summary of the Proposed Framework. In this subsection, we present a high-level summary of the proposed differential connectivity analysis (DCA) framework. Details are provided in the following subsections.

For any node \( j \in V \), any of its neighbors in any of the two networks \( k \in ne^I_j \cup ne^II_j \) must belong to one of two categories:

i) \( k \) is a common neighbor of \( j \), i.e., \( k \in ne^I_j \cap ne^II_j \equiv ne^0_j \);

ii) \( k \) is a neighbor of \( j \) in one and only one of the two networks, i.e., \( k \in ne^I_j \triangle ne^II_j \), where \( \triangle \) is the symmetric difference operator;

However, \( ne^I_j = ne^II_j \) if and only if \( ne^I_j \triangle ne^II_j = \emptyset \). Thus, if there exists a node \( k \) such that \( k \in ne^I_j \triangle ne^II_j \), then \( ne^I_j \neq ne^II_j \). Therefore, to test \( H^*_{0,j} : ne^I_j = ne^II_j \), we propose to examining whether there exists a node \( k \) such that \( k \in ne^I_j \triangle ne^II_j \equiv (ne^I_j \cup ne^II_j \setminus (ne^I_j \cap ne^II_j)) \). Specifically, for a \( k \neq j \) such that \( k \notin ne^I_j \cap ne^II_j \equiv ne^0_j \), we check whether \( k \in ne^I_j \cup ne^II_j \).

In practice, we do not observe \( ne^0_j \) and need to estimate it. Our hypothesis testing framework thus consists of two steps:

1. Estimation: We estimate the common neighborhood of each node \( j \) in the two networks, \( ne^0_j \equiv ne^I_j \cap ne^II_j \); this estimate is denoted by \( \hat{ne}^0_j \).

2. Hypothesis Testing: We test whether there exists a \( k \notin \hat{ne}^0_j \) such that \( k \in ne^I_j \cup ne^II_j \).
Details of the above two steps are described in the next two subsections, where it becomes clear that the procedure can be naturally extended to test differential connectivity in more than two networks.

From the discussion in the following subsections, it will also become evident that the estimated common neighborhood \( \hat{\text{ne}}_j^0 \) plays an important role in the validity and power of the proposed framework. In Section 3.2, we show that in order for \( \hat{\text{ne}}_j^0 \) to be useful in the hypothesis testing step, it needs to satisfy

\[
\lim_{n^I, n^{II} \to \infty} \Pr[\hat{\text{ne}}_j^0 \supseteq \text{ne}_j^0] = 1.
\]

On the other hand, as we show in Section 3.5, if the cardinality of \( \hat{\text{ne}}_j^0 \) grows large compared to that of \( \text{ne}_j^0 \), the power of the proposed framework deteriorates. In Section 3.3, we also discuss how the randomness in estimating \( \hat{\text{ne}}_j^0 \) may affect the results of the hypothesis testing step and how valid inferences can be obtained.

3.2. Estimating Common Neighbors. Given a \( j \in \mathcal{V} \), the first step of DCA involves obtaining an estimate \( \hat{\text{ne}}_j^0 \) of \( \text{ne}_j^0 \). We do not need \( \hat{\text{ne}}_j^0 \) to be a consistent estimate of \( \text{ne}_j^0 \), which usually requires stringent conditions (see, e.g., Meinshausen and Bühlmann, 2006; Zhao and Yu, 2006). Instead, we observe that under the null hypothesis \( H_{0,j}^* : \text{ne}_j^I = \text{ne}_j^{II} \), we have \( \text{ne}_j^I \cup \text{ne}_j^{II} = \hat{\text{ne}}_j^0 \). Hence, if \( \hat{\text{ne}}_j^0 \supseteq \text{ne}_j^0 \), then \( \hat{\text{ne}}_j^0 \supseteq \text{ne}_j^I \cup \text{ne}_j^{II} \), and there should be no \( k \notin \hat{\text{ne}}_j^0 \) such that \( k \in \text{ne}_j^I \cup \text{ne}_j^{II} \). Thus, we propose to test \( H_{0,j}^* \) by examining whether there exists a \( k \notin \hat{\text{ne}}_j^0 \) such that \( k \in \text{ne}_j^I \cup \text{ne}_j^{II} \). Based on the above observation, we require that

\[
\lim_{n^I, n^{II} \to \infty} \Pr[\hat{\text{ne}}_j^0 \supseteq \text{ne}_j^0] = 1.
\]

We call (5) the coverage property of estimated common neighbors.

Let \( \mathbf{X}^I \) and \( \mathbf{X}^{II} \) be two Gaussian datasets of size \( n^I \times p \) and \( n^{II} \times p \) containing measurements of the same set of variables \( \mathcal{V} \) (with \( p = |\mathcal{V}| \)) in populations I and II, respectively. Note that the data may be high-dimensional, i.e., \( p \gg \max\{n^I, n^{II}\} \). To estimate the common neighborhood \( \text{ne}_j^0 \), for \( m \in \{I, II\} \), we write

\[
x_j^m = \mathbf{X}_j^m \beta_j^m + \epsilon_j^m,
\]

where \( \beta_j^m \) is a \((p - 1)\)-vector of coefficients and \( \epsilon_j^m \) is an \( n^m \)-vector of random errors. By Gaussianity, \( \beta_j^m \neq 0 \) if and only if \( \Omega_j^m \neq 0 \), which, as discussed before, is equivalent to \( k \in \text{ne}_j^m \). Therefore, the common neighbors of node \( j \) in the two populations are

\[
\text{ne}_j^0 = \text{ne}_j^I \cap \text{ne}_j^{II} = \{k : \beta_k^I \neq 0 \& \beta_k^{II} \neq 0\}.
\]
Based on (7), an estimate of \( ne^0_j \) may be obtained from the estimated supports of \( \beta^I_{j} \) and \( \beta^{II}_{j} \).

Various procedures can be used to estimate \( \beta^I_{j} \) and \( \beta^{II}_{j} \) and, in turn, \( ne^0_j \). We present a lasso-based estimator as an example in Section 3.4. Proposition 3.2 shows that under appropriate conditions, the lasso-based estimate satisfies the coverage property (5), and is thus valid for the estimation step of DCA.

3.3. Testing Differential Connectivity. Recall that with \( \hat{ne}^0_j \supseteq ne^0_j \), if there exists a \( k \notin \hat{ne}^0_j \) such that \( k \in ne^I_j \cup ne^{II}_j \), then \( ne^I_j \neq ne^{II}_j \). As mentioned above, in GGMs, \( k \in ne^I_j \cup ne^{II}_j \) if and only if \( \beta^I_{k,j} \neq 0 \) or \( \beta^{II}_{k,j} \neq 0 \). Thus, to determine whether there exists a \( k \notin \hat{ne}^0_j \) such that \( k \in ne^I_j \cup ne^{II}_j \), we test the following hypotheses

\[
(8) H_{0,j} : \beta^I_{k,j} = 0 \quad \text{and} \quad \beta^{II}_{k,j} = 0, \quad \forall k \notin \hat{ne}^0_j \cup \{j\}.
\]

Using the Šidák correction to control false positive rate of \( H_{0,j} \) at level \( \alpha > 0 \), we control false positive rates of \( H^I_{0,j} : \beta^I_{k,j} = 0 \) and \( H^{II}_{0,j} : \beta^{II}_{k,j} = 0 \) at the level \( 1 - \sqrt{1 - \alpha} \) for all \( k \notin \hat{ne}^0_j \cup \{j\} \). We will discuss later in this subsection how to test \( H^I_{0,j} \) and \( H^{II}_{0,j} \).

The proposal outlined so far faces an important obstacle: even when \( \hat{ne}^0_j \supseteq ne^0_j \), the hypotheses \( H^I_{0,j} \) and \( H^{II}_{0,j} \) depend on the data through \( \hat{ne}^0_j \), which is a random quantity. This dependence complicates hypothesis testing: under the current procedure, we are effectively looking at the same data twice, once to formulate hypotheses and once to test the formulated hypotheses. This kind of double-peeking would usually render standard hypothesis testing procedures invalid (see, e.g., Leeb and Pötscher, 2008).

To overcome the above difficulty, we offer two different strategies. In the first, we apply sampling splitting to avoid looking at the data twice (see, e.g., Wasserman and Roeder, 2009; Meinshausen, Meier and Bühlmann, 2009). In this approach, the data are divided in two parts; the first part is used to estimate \( \hat{ne}^0_j \) and the second to test \( H_{0,j} \). The second strategy is provided in Proposition 3.3 in Section 3.4, which shows that although \( \hat{ne}^0_j \) is in general random, under appropriate conditions, the lasso-based estimate of \( \hat{ne}^0_j \) discussed in Section 3.4 converges in probability to a deterministic set, which is not affected by the randomness of the data. Thus, under those conditions, asymptotically, we can treat \( \hat{ne}^0_j \) as deterministic, and hence treat \( H^I_{0,j} \) and \( H^{II}_{0,j} \) as classical non-data-dependent hypotheses. We compare the performance of these two strategies using simulation studies, presented in Section 5.
To test $H_{0,j}^I : \beta_k^{1,j} = 0, \forall k \notin \hat{ne}_j^0 \cup \{j\}$ and $H_{0,j}^{II} : \beta_k^{II,j} = 0, \forall k \notin \hat{ne}_j^0 \cup \{j\}$, we can use recent proposals for testing coefficients in high-dimensional linear regression (e.g., Javanmard and Montanari, 2014; Zhang and Zhang, 2014; van de Geer et al., 2014; Zhao and Shojaie, 2016; Ning and Liu, 2017). To control false positive rates of $H_{0,j}^I$ and $H_{0,j}^{II}$, we need to control the family-wise error rate (FWER) on individual regression coefficients using, e.g., the Holm procedure (Holm, 1979). Alternatively, $H_{0,j}^I$ and $H_{0,j}^{II}$ can be tested using group hypothesis testing procedures that examine a group of regression coefficients, such as the least-squares kernel machines (LSKM) test (Liu, Lin and Ghosh, 2007). Although such group hypothesis testing approaches cannot be used to infer which specific edges show differential connectivity, they often result in advantages in computation and statistical power for testing $H_{0,j}^*$. An empirical comparison between the two classes to hypothesis testing strategies is presented in Section 5.

Because edges with different dependency relationship in two networks must also have different strength of connectivity, in practice, we can first apply methods described in Section 2 to find edges that show different strengths of connectivity in two networks. Then, restricted to edges that are found to have different connectivity strength, we can apply DCA to find edges that are differentially connected. Such a procedure may deliver improved power and false positive rate in ultra-high dimensional settings.

To summarize, DCA consists of two steps: estimation and hypothesis testing. These steps do not require specific methods. Rather, for the estimation step, we require that for each $j \in V$, the estimated common neighborhood, $\hat{ne}_j^0$, satisfies the coverage property $\lim_{n \to \infty} \Pr[\hat{ne}_j^0 \supseteq ne_j^0] = 1$. Moreover, we require that either $\hat{ne}_j^0$ is deterministic with high probability through, e.g., Proposition 3.3, or that the dependence between the estimation and hypothesis testing steps is severed by sample-splitting. For the hypothesis testing step, any valid high-dimensional hypothesis testing method that examines individual regression coefficients or a group of them is suitable. We arrive at the following theorem.

**Theorem 3.1.** Suppose the procedure used in the estimation step of DCA satisfies the following conditions for each $j \in V$:

1. The estimated common neighborhood of node $j$, $\hat{ne}_j^0$, satisfies the coverage property, i.e., $\lim_{n \to \infty} \Pr[\hat{ne}_j^0 \supseteq ne_j^0] = 1$;
2. Either the estimated common neighborhood $\hat{ne}_j^0$ is deterministic with probability tending to one, or the data used to test hypotheses $H_{0,j}^m$ for $m \in \{I, II\}$ are independent of the data used to estimate $\hat{ne}_j^0$. 
Then, if for \( m \in \{I, II\} \) the hypothesis testing procedure for testing \( H_{0,m}^j : \beta_{\neq j}^n = 0 \) is asymptotically valid, DCA asymptotically controls the false positive rate of \( H_{0,j}^j : ne_j^I = ne_j^II \).

Theorem 3.1 outlines a general framework that can incorporate many estimation and inference procedures. Moreover, with appropriate high-dimensional estimation and inference procedures, Theorem 3.1 is applicable in the high-dimensional setting. In Section 3.4, we discuss two alternative strategies based on lasso that satisfy the requirements of Theorem 3.1.

3.4. Using Lasso for DCA. Lasso neighborhood selection (Meinshausen and Bühlmann, 2006) is a convenient procedure for estimating \( ne_j^0 \). We show that lasso is a valid procedure for the estimation step in DCA. However, lasso is not the only valid estimation procedure: any procedure that satisfies the requirements of Theorem 3.1 is valid. In Section 3.5, we discuss the power of DCA with lasso as the estimation procedure. There, we also present a sufficient condition for the DCA to asymptotically achieve perfect power.

In this section, we present two propositions regarding lasso neighborhood selection, which show that under appropriate conditions, the estimate \( \hat{ne}_j^0 \) satisfies the coverage property, \( \lim_{n^I, n^II \to \infty} \Pr[\hat{ne}_j^0 \supseteq ne_j^0] = 1 \), and is deterministic with high probability. Together, these results imply that lasso neighborhood selection satisfies the requirements of Theorem 3.1 both with and without sampling splitting.

We now present Propositions 3.2 and 3.3. As mentioned in Section 3.1, Proposition 3.2 implies that lasso neighborhood selection is a valid method for estimation in our framework, and Proposition 3.3 relieves us from using sample-splitting to circumvent double-peeking by our procedure. The conditions are summarized below. Note that we only present lasso here as an example—other methods can be incorporated into DCA so long as they satisfy the requirements of Theorem 3.1. A number of methods that fit into the DCA framework will be numerically evaluated in Section 5.

Next, we state the sufficient conditions for our propositions.

(A1) For \( m \in \{I, II\} \), rows of the data \( X^m \) are independent and identically distributed Gaussian random vectors: \( X^m \sim_{i.i.d.} N_p(0, \Sigma^m) \), where, without loss of generality, we assume \( \text{diag}(\Sigma^m) = 1 \). Further, the minimum and maximum eigenvalues of \( \Sigma^m \) satisfy
\[
\liminf_{n^m \to \infty} \phi_{\min}^2 (\Sigma^m) > 0 \quad \text{and} \quad \limsup_{n^m \to \infty} \phi_{\max}^2 (\Sigma^m) < \infty.
\]

(A2) For \( m \in \{I, II\} \) and any variable \( j \in V \), the sample size \( n^m \), dimension \( p \), lasso neighborhood selection tuning parameters \( \lambda_j^m \), number
of neighbors \( q^m_j \equiv |\text{ne}_m^j| \), and minimum non-zero coefficients \( b_{\text{min}}^{m,j} \equiv \min\{ |\beta_k^{m,j}| : \beta_k^{m,j} \neq 0 \} \), where \( \beta^{m,j} \) is defined in (6), satisfy

\[
\limsup_{n^m \to \infty} \lambda^m_j q^m_j < \infty, \quad \lim_{n^m \to \infty} \frac{\sqrt{\log p}}{n^m} q^m_j = 0, \quad \lim_{n^m \to \infty} \frac{\lambda^m_j \sqrt{q^m_j}}{b_{\text{min}}^{m,j}} = 0.
\]

(A3) For \( m \in \{I, II\} \) and any variable \( j \in \mathcal{V} \), define the sub-gradient \( \hat{\tau}^{m,j} \) based on the stationary condition of (12)

\[
\hat{\tau}^{m,j} = \frac{1}{n \lambda^m_j} \mathbb{E} \left[ X_{\neq m}^j \left( x_j^m - X_{\neq m}^j \hat{\beta}^{m,j} \right) \right].
\]

We assume \( \hat{\tau}^{m,j} \) satisfies \( \limsup_{n^m \to \infty} \left\| \hat{\tau}^{m,j}_{\neq m} \right\|_\infty \leq 1 - \delta^m \) such as

\[
\lim_{n^m \to \infty} \sqrt{\frac{\log p}{n^m}} \frac{q_j}{\lambda^m_j \delta^m} = 0,
\]

and

\[
\lim_{n^m \to \infty} \frac{q_j^m}{\lambda^m_j} \sqrt{\frac{\log p}{n^m}} \left( \min_{k \in \text{ne}_m^j \setminus \text{ne}_m^j} \left| \Sigma_{(\text{ne}_m^j \setminus \text{ne}_m^j)}^{m,j} \right|^{-1} \hat{\tau}^{m,j}_{\neq m} \right)_k^{-1} = 0.
\]

Condition (A1) characterizes the data distribution. Combining the first two requirements of (A2), for \( m \in \{I, II\} \), we get \( q_j^m = \mathcal{O}((n^m / \log(p))^{1/4}) \). The third constraint in (A2) is the \( \beta\)-min condition, which prevents the signal from being too weak to be detected; this condition may be relaxed to allow the presence of some weak signal variables (Zhao, Witten and Shojaie, 2021). Note that although our goal is to test the difference in connectivity in two networks, Condition (A2) does not require the difference in signal strength to be large under the null hypothesis. In additional, (A2) requires the tuning parameters \( \lambda^m_j \) to approach zero at a slower rate than \( q_j^m \sqrt{\log(p)}/n^m \), which is the minimum tuning parameter rate for prediction consistency of lasso with Gaussian data (see, e.g., Bickel, Ritov and Tsybakov, 2009). Since \( \lim_{n^m \to \infty} q_j^m \sqrt{\log(p)}/n^m / \lambda^m_j = 0 \) by (A2), condition (A3) requires that the tuning parameter \( \lambda \) does not converge to any transition points too fast, where some entries of \( \hat{\beta}^{m,j} \) change from zero to nonzero, or vice versa. (A3) also requires that

\[
\min_{k \in \text{ne}_m^j \setminus \text{ne}_m^j} \left| \Sigma_{(\text{ne}_m^j \setminus \text{ne}_m^j)}^{m,j} \right|^{-1} \hat{\tau}^{m,j}_{\neq m} \right)_k
\]

does not converge to zero too fast.
Proposition 3.2. Suppose conditions (A1) and (A2) hold for variable $j \in V$. Then $\hat{ne}_j^0$ estimated using lasso neighborhood selection satisfies

$$\lim_{n^I, n^II \to \infty} \Pr \left[ \hat{ne}_j^0 \supseteq ne_j^0 \right] = 1.$$  

(10)

To establish that lasso-based estimates of neighborhoods are deterministic with high probability, in Proposition 3.3 we establish a novel relationship between the lasso neighborhood selection estimator,

$$\hat{\beta}^{m,j} \equiv \arg \min_{b \in \mathbb{R}^{p-1}} \left\{ \frac{1}{2n} \left\| x_j^m - X_{\setminus j}^m b \right\|_2^2 + \lambda^m_j \| b \|_1 \right\},$$  

(11)

and its noiseless (and hence deterministic) counterpart

$$\tilde{\beta}^{m,j} \equiv \arg \min_{b \in \mathbb{R}^{p-1}} \left\{ \mathbb{E} \left[ \frac{1}{2n} \left\| x_j^m - X_{\setminus j}^m b \right\|_2^2 \right] + \lambda^m_j \| b \|_1 \right\}.$$  

(12)

Proposition 3.3. Suppose conditions (A1) – (A3) hold for variable $j \in V$. Then $\hat{ne}_j^0$ estimated using lasso neighborhood selection satisfies

$$\lim_{n^I, n^II \to \infty} \Pr \left[ \hat{ne}_j^0 = \tilde{ne}_j^0 \right] = 1,$$

(13)

where $\tilde{ne}_j^0 \equiv \text{supp}(\tilde{\beta}^{I,j}) \cap \text{supp}(\tilde{\beta}^{II,j})$, and $\tilde{\beta}^{I,j}$ and $\tilde{\beta}^{II,j}$ are defined in (12).

Propositions 3.2 and 3.3 are proven in Sections S1 and S2 of the Supplementary Material, respectively. Similar results for deterministic design matrices were proposed in Zhao, Witten and Shojaie (2021). The result in Proposition 3.3 should not be confused with the variable selection consistency of lasso (Meinshausen and Bühlmann, 2006), which shows that under the stringent irrepresentability condition, the selected neighborhoods converge in probability to the true neighborhoods, i.e., $\lim_{n^I, n^II \to \infty} \Pr \left[ \hat{ne}_j^0 = ne_j^0 \right] = 1$. Proposition 3.3 only shows that the selected neighborhoods converge to deterministic sets, $\tilde{ne}_j^0$, and requires milder conditions.

3.5. Power of DCA with Lasso in the Estimation Step. In Section 3.2, we argued that the estimated common neighborhood $\hat{ne}_j^0$ needed to satisfy the coverage property, i.e., $\lim_{n^I, n^II \to \infty} \Pr[\hat{ne}_j^0 \supseteq ne_j^0] = 1$. In this section, we discuss how the cardinality of $\hat{ne}_j^0$ affects the power of DCA. We also discuss a sufficient condition, where, using lasso in the estimation step, DCA could achieve asymptotically perfect power for detecting differential connectivity.
As mentioned in Section 3.3, to examine $H_{0,j} : ne_j^I \neq ne_j^H$, in the second step of DCA, we test whether there exists a $k \notin \hat{ne}_j^0$ such that $k \in ne_j^1 \cup ne_j^H$. If the estimated neighborhood of variable $j$ is too large, such that $\hat{ne}_j^0 \supseteq ne_j^1 \cup ne_j^H$, then for any $k \notin \hat{ne}_j^0 \cup \{j\}$, $x_j^m \perp x_k^m \mid x_j \setminus \{j,k\}$ for $m \in \{I, II\}$. In this case, we will not be able to identify differential connectivity of node $j$. Thus, even though the validity of DCA requires that $\hat{ne}_j^0$ achieves the coverage property, $\hat{ne}_j^0$ should not be exceedingly larger than $ne_j^0$.

To examine the power of DCA with the lasso-based estimate of joint neighbor, $\hat{ne}_j^0$, suppose $|ne_j^H| = o(|ne_j^I|)$. Belloni and Chernozhukov (2013) show that, under mild conditions, with high probability, $|ne_j^m| \asymp |\hat{ne}_j^m|$, where $\asymp$ denotes that two quantities are of the same asymptotic order. Therefore, with high probability, $|\hat{ne}_j^0| \leq |\hat{ne}_j^H| \asymp |ne_j^H| = o(|ne_j^I|)$, and hence, $\hat{ne}_j^0 \neq ne_j^I$. Similarly, if $|ne_j^I| = o(|ne_j^H|)$, then with high probability $\hat{ne}_j^0 \neq ne_j^H$. Thus, if $|ne_j^I|$ and $|ne_j^H|$ are not of the same asymptotic order, then, with high probability, $\hat{ne}_j^0 \neq ne_j^1 \cup ne_j^H$. Given the coverage property, $\hat{ne}_j^0 \supseteq ne_j^1 \cap ne_j^H$, if $|ne_j^1|$ and $|ne_j^H|$ are not of the same order, the above reasoning implies that $\hat{ne}_j^0 \neq ne_j^1 \triangle ne_j^H$ with high probability. In this case, with any conditional testing method that achieves asymptotic power one, i.e., those for which the test statistic $t$ satisfies, for any $\alpha > 0$, $\lim_{n \to \infty} \Pr(t < \alpha) = 1$ under the alternative hypothesis, DCA is asymptotically guaranteed to detect the differential connectivity of $j$.

Note that although $|ne_j^H| = o(|ne_j^I|)$ implies that $|ne_j^1 \cap ne_j^H| = o(|ne_j^1 \cup ne_j^H|)$, the assumption $|ne_j^1 \cap ne_j^H| = o(|ne_j^1 \cup ne_j^H|)$ alone is insufficient for ensuring that power converges to one. This is because if $|ne_j^I| \asymp |ne_j^H|$, although practically unlikely, because $|ne_j^I \cap ne_j^H|$ is much smaller than $|ne_j^1 \cup ne_j^H|$, it is still theoretically possible that $\hat{ne}_j^0 \supseteq ne_j^1 \cup ne_j^H$ and $\hat{ne}_j^0 \supseteq ne_j^1 \cup ne_j^H$, in which case $\hat{ne}_j^0 \supseteq ne_j^1 \cup ne_j^H$ and it is impossible to reject the null hypothesis. On the other hand, if we assume the more stringent neighborhood stability condition (Meinshausen and Bühlmann, 2006), then with high probability, $\hat{ne}_j^0 \neq ne_j^1 \cup ne_j^H$ if $|ne_j^I \cap ne_j^H| = o(|ne_j^1 \cup ne_j^H|)$. Thus, with this assumption, DCA is guaranteed to have asymptotically perfect power even if $|ne_j^H| = o(|ne_j^I|)$ does not hold. Note, however, that neighborhood stability is not needed for the validity of DCA, and we only discussed this alternative assumption to analyze the power of DCA.

The conditions presented in the above special case are sufficient and not necessary. Nonetheless, the scenario sheds light on the power properties of DCA. We defer to future research a more thorough assessment of power properties of DCA.
4. Differential Connectivity in Breast Cancer Subtypes. In this section, we revisit the TCGA breast cancer data briefly discussed in Section 1 to further demonstrate the utility of DCA for detecting differential connectivity in genetic networks.

Breast cancer has multiple clinically verified subtypes (Perou et al., 2000) that have been shown to have distinct prognostics (Jönsson et al., 2010). Based on the expression of estrogen receptor (ER), breast cancer can be classified into ER positive (ER+) and ER negative (ER-) subtypes. ER+ breast cancer has a larger number of estrogen receptors, and has better survival prognosis than ER- breast cancer (Carey et al., 2006). The genetic pathways of ER+ and ER- subtypes are expected to be similar, but also show some important differences. Understanding such differences could be critical to help researchers better understand breast cancer. To investigate differences in genetic pathways between ER+ and ER- breast cancer patients, we obtain gene expression data from the Cancer Genome Atlas (TCGA). The data contain the expression levels of $p = 358$ genes in cancer related pathways from KEGG for $n^I = 117$ ER- and $n^{II} = 407$ ER+ breast cancer patients.
Following the findings of the simulation study in Section 5, we use the GraceI test after naïve (i.e., no sample-splitting) lasso neighborhood selection in this analysis to examine the difference in genetic pathways between ER+ and ER- breast cancer patients. Controlling the family-wise error rate at $\alpha = 0.1$ level using Holm’s procedure, we find 60 differentially connected genes using this procedure.

As a comparison, we performed quantitative test of Gill, Datta and Datta (2014), which detected that 75 out of 358 genes in the dataset are differentially connected in two networks at family-wise error rate level of $\alpha = 0.1$.

Differentially connected edges using DCA are shown in Figure 3. Genes identified as differentially connected are also scientifically relevant. Specifically, all of the genes that have at least three differential connections identified by DCA have already been found by previous research to be associated with the subtype, progression and prognostics of breast cancer. These highly differentially connected genes are: laminin subunit $\beta_1$ (LAMB1), matrix Metalloproteinase-2 (MMP2), platelet-derived growth factor receptor $\alpha$ (PDGFRA), phosphoinositide 3’-kinases $\delta$ (PIK3CD), runt-related transcription factor 1 (RUNX1T1) and TGF-$\beta$ receptor type-2 (TGFBR2). For instance, LAMB1 has been found to be an informative prognostic biomarker for ER- and ER+ tumors. In particular, Pellegrini et al. (1995) find laminin expression to be associated with a good prognosis in receptor-negative tumors, but worst prognosis in receptor-positive tumors. In particular, Pellegrini et al. (1995) find laminin expression to be associated with a good prognosis in receptor-negative tumors, but worst prognosis in receptor-positive tumors. Similarly, Jeziorska and Motyl (2009) lament that activation of MMP2 is “considered a very sensitive indicator of cancer metastasis.” Moreover, observing that lower levels of MMP2 expression are linked to favorable prognosis in patients with a hormone receptor-negative tumor, the authors emphasize to the therapeutic potential of this gene. In another study, Carvalho et al. (2005) report that over-expression of PDGFRA is associated with tumor progression in breast cancer. Moreover, the authors find that PDGFRA expression is associated with breast carcinomas and lymph node metastasis. Interestingly, they also found that observed mutations in this gene were not correlated with higher expression levels, which may further suggest that gene regulatory networks play a role in modulating the expression of this gene.

The other genes, namely PIK3CD, RUNX1T1 and TGFBR2, are also supported by the existing literature. In particular, Sawyer et al. (2003) report that PIK3CD can regulate cell migration in breast cancer. Similarly, pointing out that RUNX1 is expressed in breast epithelia and may be misregulated during tumorigenesis, Janes (2011) argues that RUNX1 could act as a tumor suppressor gene in breast cancer. However, the in vitro experiments in this paper indicate that hyper-proliferation in RUNX1-deficient
breast epithelia relies on other transcription factors, which also emphasize the role of gene regulatory networks. Finally, Ma et al. (2012) and Busch et al. (2015) provide evidence to support the role of TGFBR2 in breast cancer. In particular, using data from a large consortium, Ma et al. (2012) find that minor allele homozygotes (AA) of one of the SNPs in TGFBR2 had significantly reduced risk of breast cancer compared to major allele carriers (AG or GG). Moreover, Busch et al. (2015) find that TGFBR2 expression regulates breast cancer cell growth and survival and emphasize the value of TGFBR2 as both a prognostic and therapeutic marker in pre-menopausal breast cancer. Using mouse model experiments, the authors also confirm that loss of TGFBR2 expression in mammary fibroblasts is linked to tumor initiation and metastasis. The above evidence, and others not listed here, confirm the potential of differential network analysis as a complementary tool for discovering prognostic and therapeutic biomarkers.

5. Simulation Studies. Motivated by the TCGA data analyzed in Section 4, in this section we present results of a simulation study that evaluates the power and false positive rate of the DCA framework. To provide a comprehensive evaluation, we use various choices of procedures in the estimation and hypothesis testing steps in DCA. However, we do not compare these powers with those of quantitative tests that examine the equality of inverse covariances. This is because, as demonstrated in Table 1, quantitative tests do not control the type-I error rate of qualitative hypotheses. Therefore, comparison of powers with these methods would not be meaningful.

We generate $\mathcal{E}^I$ from a power-law degree distribution with power parameter 5, $|\gamma| \equiv p = 200$ and $|\mathcal{E}^I| = p(p-1)/100$; this corresponds to an edge density of 0.02 in graph $G^I$. Power-law degree distributions are able to produce graphs with hubs, which are expected in real-world networks (Newman, 2003). To simulate $\mathcal{E}^{II}$, among the 100 most connected nodes in $G^I$, we randomly select 20 nodes, remove all the edges that are connected to them, and then randomly add edges to graph $G^{II}$ so that $|\mathcal{E}^{II}| = |\mathcal{E}^I|$. To simulate $\Omega^I$, for $j \neq k$, we let

$$\Omega^I_{jk} = \begin{cases} 
0 & (j,k) \notin \mathcal{E}^I \\
0.5 & (j,k) \in \mathcal{E}^I, \text{ with 50% probability} \\
-0.5 & (j,k) \in \mathcal{E}^I, \text{ with 50% probability}
\end{cases}$$
To simulate $\Omega^\Pi$, for $j \neq k$, we let

$$\Omega^\Pi_{jk} = \begin{cases} 
\Omega^I_{jk} & (j, k) \in E^I \cap E^\Pi \\
0 & (j, k) \notin E^\Pi \\
0.5 & (j, k) \in E^\Pi \setminus E^I, \text{ with 50\% probability} \\
-0.5 & (j, k) \in E^\Pi \setminus E^I, \text{ with 50\% probability}
\end{cases}$$

Finally, for $m \in \{I, \Pi\}$, we let $\Omega^m_{jj} = \sum_{k \neq j} |\Omega^m_{jk}| + u^m$ for $j = 1, \ldots, p$, where $u^m$ is chosen such that $\phi^2_{\min}(\Omega^m) = 0.1$, where $\phi^2_{\min}(\Omega^m)$ is the smallest eigenvalues of $\Omega^m$. We then normalize $\Omega^m$ such that the diagonal entries of $\Sigma^m$ are one. Figure S1 in Section S3 in the Supplementary Materials shows the distribution of non-zero inverse covariances in $\Omega^I$ and $\Omega^\Pi$. From $\Omega^I$ and $\Omega^\Pi$, we generate $X^I \sim \text{i.i.d.} \ N_p(0, [\Omega^I]^{-1})$ and $X^\Pi \sim \text{i.i.d.} \ N_p(0, [\Omega^\Pi]^{-1})$, where $n^I = n^\Pi = n \in \{100, 200, 400, 800\}$.

To estimate common neighbors of each node $j \in V$, we use lasso neighborhood selection, with tuning parameters chosen by 10-fold cross-validation (CV). We either use sample-splitting to address the issue of double-peeking, with half of samples used to estimate $\hat{\nu}^0_j$ and the other half to test $H^0_{0,j}$, or use a naïve approach, where the whole dataset is used to estimate $\hat{\nu}^0_j$ and to test $H^0_{0,j}$: $\nu^I_j = \nu^\Pi_j$; the latter approach is justified by Proposition 3.3. To examine $H^0_{0,j}$ for each $j = 1, \ldots, p$, we consider LSKM (Liu, Lin and Ghosh, 2007), which is a group hypothesis testing methods, and the GraceI test (Zhao and Shojaie, 2016), which examines individual regression coefficients. As a result, we compare in total 4 approaches: \{naïve lasso neighborhood selection, sample-splitting lasso neighborhood selection\} × \{LSKM, GraceI\}.

Note that similar to the discussion in Meinshausen and Bühlmann (2006), it is possible that a node-pair $(j, k)$ is identified to be differentially connected in testing $H^*_0_{0,j}: \nu^I_j = \nu^\Pi_j$, but not so in testing $H^*_0_{0,k}: \nu^I_k = \nu^\Pi_k$. To mitigate this issue, we used the “OR” rule (Meinshausen and Bühlmann, 2006) in the simulation studies and the cancer genetics application presented in Section 4. With the “OR” rule, an edge becomes a false positive if it is a false positive in any of the two node-wise tests. Hence, we should control the false positive rate at level $\alpha/2$ for the node-wise tests (i.e., Bonferroni correction). Based on a similar reasoning, the “AND” rule is also valid if the node-wise tests are controlled at level $\alpha$.

Figure 4 shows average false positive rates of falsely rejecting $H^*_0_{0,j}: \nu^I_j = \nu^\Pi_j$, as well as average power of various DCA variants based on $R = 100$ repetitions. Let $z^r_{j,r}$ be the decision function based on the GraceI test or LSKM: $z^r_{j,r} = 1$ if $H^*_0_{0,j}: \nu^I_j = \nu^\Pi_j$ is rejected in the $r$th repetition, and
The average false positive rate and power of rejecting $H_{0,j}^*$. The axis for false positive rate is on the left of each panel, whereas the axis for power is on the right.

\[ \text{TIER} = \frac{\sum_{r=1}^{R} \left\{ \sum_{j \in \mathcal{V}: ne_{j,r}^I = ne_{j,r}^II} z_{j,r} \right\}}{\sum_{r=1}^{R} \left\{ j \in \mathcal{V} : ne_{j,r}^I = ne_{j,r}^II \right\}} , \]

i.e., the proportion of null hypotheses in $R$ repetitions that we falsely reject $H_{0,j}^*$. For $t \in \{1, 3, 5, 10\}$, the average power of rejecting $H_{0,j}^* : ne_{j}^I = ne_{j}^II$ when $ne_{j}^I$ and $ne_{j}^II$ differ by at least $t$ members is defined as

\[ \text{Pt} = \frac{\sum_{r=1}^{R} \left\{ \sum_{j \in \mathcal{V}: |ne_{j,r}^I \triangle ne_{j,r}^II| \geq t} z_{j,r} \right\}}{\sum_{r=1}^{R} \left\{ j \in \mathcal{V} : |ne_{j,r}^I \triangle ne_{j,r}^II| \geq t \right\}} , \]

where “$\triangle$” denotes the symmetric difference of two sets.

The simulation reveals several interesting patterns. First, naïve procedures which use the same data to estimate $\hat{ne}_{j}^0$ and test $H_{0,j}^m$, $m \in \{I, II\}$ tend to have better statistical power than their sample-splitting counterparts. This is understandable, as sample-splitting only uses half of the data for hypothesis testing. More surprisingly, naïve procedures also better control the false positive rate than sample-splitting procedures. This is because $z_{j,r} = 0$ otherwise.
the event $\tilde{\eta}_j^0 \supseteq \eta_j^0$, which is crucial for controlling the false positive rate and is guaranteed to happen with high probability asymptotically, is less likely to happen with the smaller samples available for the sample-splitting estimator. In addition, we can see that LSKM has better power than the GraceI test for smaller sample sizes (LSKM also has a slightly worse control of the false positive rate than GraceI). But as sample size increases, the power of GraceI eventually surpasses LSKM. Finally, as expected, the probability of rejecting $H_{0,j}^*: \eta_j^I = \eta_j^II$ is higher when $\eta_j^I$ and $\eta_j^II$ differ by more elements.

6. Conclusion. In this paper, we highlighted challenges of identifying differential connectivity in high-dimensional networks using existing approaches, and proposed a new hypothesis testing framework, called differential connectivity analysis (DCA), for identifying differences in two networks. The proposed method is implemented in the DCA R-package available on GitHub (https://github.com/sen-zhao/DCA).

DCA can incorporate various estimation and hypothesis testing methods, and can be easily extended to test for differential connectivity in multiple networks. We discussed two options for estimation and inference: sample-splitting, which breaks down the dependence between estimation and hypothesis testing, and naïve inference, which utilizes the fact that the estimated support of lasso is deterministic with high probability. Besides these approaches, another option is to build on recent advances in conditional hypothesis procedures (see, e.g., Lee et al., 2016; Tibshirani et al., 2016). We leave to future work the exploration of whether conditional hypothesis testing procedures can be adapted to fit into DCA. Exploring the feasibility of incorporating non-convex estimation methods (e.g., Fan and Li, 2001; Zhang, 2010) in DCA could also be a fruitful area of research.

There are two possible approaches for speeding up the computation and reducing the number of tests in DCA. First, DCA involves running $2p$ Lasso neighborhood selection, which can be run in parallel to speed up the computation. Second, instead of performing $O(p^2)$ hypothesis tests, we can perform $p$ group-based hypothesis tests (Liu, Lin and Ghosh, 2007), as discussed in Section 3.3. This approach is appropriate for settings where the goal is to identify differentially connected nodes. However, with group-based hypothesis testing procedures, we are no longer able to pinpoint the specific edges that are differentially connected. If this is not a concern, then group-based tests can be used to improve the computation and reduce the number of hypotheses.

References.


JÖNSSON, G., STAFA, J., VALLON-CHRISTERSSON, J., RINGNÉR, M., HOLM, K.,


Sawyer, C., Sturje, J., Bennett, D. C., O’Hare, M. J., Allen, W. E., Bain, J., etc.


