

Chapter 2

Approximately Factorizable Networks

Abstract In factorizable networks, the adjacency (connection strength) between two nodes can be factored into node-specific contributions, named node “conformity”. Often the i th node conformity, CF_i is approximately equal to the scaled connectivity $k_i/\text{sum}(k)$ of the i th node. We describe (a) an algorithm for computing the conformity CF and for measuring the “factorizability” of a general network, and (b) a module- and CF-based decomposition of a general adjacency matrix, which can be used to arrive at a parsimonious description of a network. Approximately factorizable networks have important practical and theoretical applications, e.g., we use them to derive relationships between network concepts. Collaborative work with **Jun Dong** has shown that network modules (i.e., subnetworks comprised of module nodes) tend to be approximately factorizable (Dong and Horvath BMC Syst Biol 1(1):24, 2007).

2.1 Exactly Factorizable Networks

We define an adjacency matrix A to be exactly factorizable if, and only if, there exists a vector CF with nonnegative elements such that

$$A_{ij} = CF_i CF_j \text{ for all } i \neq j. \quad (2.1)$$

Below, we show that CF is uniquely defined if the network contains $n \geq 3$ nodes and if $A_{ij} > 0$. We call CF_i the **conformity** of node i . While the term “factorizable network” was first proposed in [Dong and Horvath \(2007\)](#), numerous examples of these types of networks can be found in the literature. A recent physical model for experimentally determined protein–protein interactions is exactly factorizable ([Deeds et al. 2006](#)). In that model, the ‘affinity’ A_{ij} between proteins i and j is the product of conformities $CF_i = \exp(-K_i)$, where K_i is the number of hydrophobic residues in the i th protein. Another related example is an exactly factorizable random network model for which the edges between pairs of nodes are drawn according to a linking probability function ([Servedio et al. 2004](#)). In multi-graph models, CF_i has been used to represent the propensity for the i th node to form

an edge (Ranola et al. 2010). One can easily show that the vector CF is not unique if an exactly factorizable network contains only $n = 2$ nodes. However, for $n > 2$ the conformity is uniquely defined when dealing with a weighted network where $A_{ij} > 0$. In an exercise, you are asked to prove the uniqueness.

2.2 Conformity for a Non-Factorizable Network

Equation (2.21) provides an explicit formula for the conformity of a weighted, exactly factorizable network. For a general, non-factorizable network, we describe here how to compute the conformity by optimizing an objective function (Dong and Horvath 2007). In the following, we assume a general $n \times n$ adjacency matrix A where $n > 2$. Let $v = (v_1, v_2, \dots, v_n)$ be a vector of length n . We define the conformity as a vector v^* that *minimizes* the following objective function:

$$f(v) = \sum_i \sum_{j \neq i} (A_{ij} - v_i v_j)^2 = \|A - (I - \text{diag}(v^2) + vv^\tau)\|_F^2,$$

where $\|\cdot\|_F$ denotes the Frobenius matrix norm (1.4). An equivalent but more convenient approach is to define the conformity CF as the maximizer of the following **factorizability function**:

$$\begin{aligned} F_A(v) &= 1 - \frac{f(v)}{f(0)} = 1 - \frac{\sum_i \sum_{j \neq i} (A_{ij} - v_i v_j)^2}{\sum_i \sum_{j \neq i} (A_{ij})^2} \\ &= 1 - \frac{\|A - (vv^\tau - \text{diag}(v^2) + I)\|_F^2}{\|A - I\|_F^2}, \\ &= 1 - \frac{\|A - A_v\|_F^2}{\|A - I\|_F^2}, \end{aligned} \quad (2.2)$$

where $A_v = vv^\tau - \text{diag}(v^2) + I$ is a matrix whose diagonal elements equal 1. One can easily show that if v^* maximizes $F_A(v)$, then $-v^*$ also maximizes $F_A(v)$. Further, all components of v^* must have the same sign, since otherwise flipping the sign of the negative components leads to a higher value of $F_A(v)$. This leads us to the following:

Definition 2.1 (Conformity, Factorizability). We define the conformity CF as the vector with nonnegative entries that maximizes $F_A(v)$. If there is more than one such maximizer, then a maximizer closest to $k/\sqrt{\text{sum}(k)}$ is chosen. Further, we define the factorizability $F(A)$ as the corresponding maximum value $F_A(CF)$.

Thus, the **network factorizability measure** is defined as follows:

$$F(A) = F_A(CF) = \max_v (F_A(v)) = 1 - \frac{\|A - A_{CF}\|_F^2}{\|A - I\|_F^2}, \quad (2.3)$$

where

$$A_{CF} = CFCF^\tau - \text{diag}(CF^2) + I.$$

A_{CF} can be interpreted as the exactly factorizable network which best approximates A (according to the Frobenius norm). In general, the closer $F(A)$ is to 1, the better A_{CF} approximates A . Note that $F_A(0) = 0$ implies that $F(A) \geq 0$.

An algorithmic definition of the factorizability $F(A)$ is provided in (2.11). We find that for most real networks, the conformity is highly related to the first eigenvector of the adjacency matrix, i.e., the conformity vector CF is roughly equal to $\sqrt{d_1}u_1$ where d_1 is the largest eigenvalue of A and u_1 is the corresponding unit length eigenvector with positive components. The idea of using a variant of the singular value decomposition for decomposing an adjacency matrix has been proposed by several authors. However, we prefer to define the conformity as a maximizer of the factorizability function $F_A(v)$ (2.2) for the following reasons: First, the factorizability satisfies that $F_A(CF) = 1$ if, and only if, A is exactly factorizable network with $A_{ij} = CF_i CF_j$. Second, we prefer to define the conformity without reference to the diagonal elements A_{ii} of the adjacency matrix. Third, the definition naturally fits within the framework of least squares factor analysis where conformity can be interpreted as the first factor (de Leeuw and Michailidis 2000). While network analysis focuses on the adjacency matrix, factor analysis takes as input a correlation or covariance matrix.

2.2.1 Algorithm for Computing the Node Conformity

This technical section may be skipped by readers who are not interested in learning the algorithmic definition of the node conformity and factorizability. We describe an algorithm for finding a vector v (with nonnegative components) that maximizes the objective function $F_A(v)$ (2.2). In general, $F_A(v)$ may have multiple maximizers as can be demonstrated with the block diagonal simulated example (3.37) by choosing $n^{(1)} = n^{(2)}$ and $b_1 = b_2$. By forming the first derivative of the factorizability function $F_A(v)$ in terms of v_i , one can show that a local maximum satisfies

$$\sum_{j \neq i} A_{ij} CF_j = CF_i \sum_{j \neq i} CF_j^2, \quad (2.4)$$

i.e.,

$$(A - I + \text{diag}(CF^2)) CF = CF \|CF\|_2^2. \quad (2.5)$$

Equation (2.5) suggests that the conformity is an eigenvector of the “hat” adjacency matrix which also depends on the conformity

$$\hat{A} = A - I + \text{diag}(CF^2).$$

Denote by $CF^{(s-1)}$ an estimate of the conformity CF. Next define

$$\hat{A}^{(s-1)} = A - I + \text{diag} \left(\left(CF^{(s-1)} \right)^2 \right). \quad (2.6)$$

Define a new estimate of the conformity by

$$CF^{(s)} = \sqrt{\hat{d}_1^{(s-1)}} \hat{u}_1^{(s-1)}, \quad (2.7)$$

where $\hat{d}_1^{(s-1)}$ and $\hat{u}_1^{(s-1)}$ denote the largest eigenvalue and corresponding unit length eigenvector of $\hat{A}^{(s-1)}$. One can easily show that all the components of $\hat{u}_1^{(s-1)}$ must have the same sign and we assume without loss of generality nonnegative components. Our algorithm for computing the conformity is monotonic as can be proven with the following:

Lemma 2.1. *If A denotes a symmetric real matrix with eigenvalues d_1, \dots, d_n sorted according to their absolute values, i.e., $|d_1| \geq |d_2| \geq \dots \geq |d_n|$, and the corresponding orthonormal eigenvectors are denoted by u_1, \dots, u_n , then $\|A - vv^\tau\|_F^2$ is minimized at $v^* = \sqrt{|d_1|}u_1$.*

The proof can be found in [Horn and Johnson \(1991\)](#).

Lemma 2.1 with $A = \hat{A}^{(s-1)}$ implies that

$$\begin{aligned} & \left\| A - I + \text{diag} \left(\left(CF^{(s-1)} \right)^2 \right) - CF^{(s-1)} \left(CF^{(s-1)} \right)^\tau \right\|_F^2 \\ & \geq \left\| A - I + \text{diag} \left(\left(CF^{(s)} \right)^2 \right) - CF^{(s)} \left(CF^{(s)} \right)^\tau \right\|_F^2. \end{aligned} \quad (2.8)$$

Considering the diagonal elements, one can easily show that

$$\begin{aligned} & \left\| A - I + \text{diag} \left(\left(CF^{(s-1)} \right)^2 \right) - CF^{(s-1)} \left(CF^{(s-1)} \right)^\tau \right\|_F^2 \\ & \geq \left\| A - I + \text{diag} \left(\left(CF^{(s)} \right)^2 \right) - CF^{(s)} \left(CF^{(s)} \right)^\tau \right\|_F^2. \end{aligned} \quad (2.9)$$

Thus, we arrive at the following:

$$F_A \left(CF^{(s)} \right) \geq F_A \left(CF^{(s-1)} \right), \quad (2.10)$$

which suggests a monotonic algorithm for computing CF. Equation (3.24) suggests to choose $k/\sqrt{\text{sum}(k)}$ as a starting value of the algorithm. These comments give rise to the following:

Definition 2.2 (Algorithmic Definition of Conformity, Factorizability). For a general network A , set $CF^{(1)} = k/\sqrt{\text{sum}(k)}$ and apply the monotonic iterative

algorithm described by (2.6) and (2.7). If the limit $CF^{(\infty)}$ exists, we define it as the conformity $CF = CF^{(\infty)}$. Further, we define the network factorizability (2.3) as:

$$F(A) = F_A(CF) = 1 - \frac{\|A - A_{CF}\|_F^2}{\|A - I\|_F^2}. \quad (2.11)$$

Note that the conformity satisfies (2.5) by definition of convergence. The monotonic algorithm described by (2.6) and (2.7) can be derived as a majorization–minimization (MM) algorithm (Lange 2004; Ranola et al. 2010), and it can be considered as a special case of an algorithm described for fitting a least squares factor analysis model with one factor (de Leeuw and Michailidis 2000; Gifi 1990).

2.3 Module-Based and Conformity-Based Approximation of a Network

A factorizable network allows for a simple and parsimonious parametrization in terms of the components of the conformity vector. But in general, $n * (n - 1)/2$ upper diagonal elements A_{ij} are needed to parameterize an adjacency matrix A . While most adjacency matrices A are *not* approximately factorizable, they may contain modules that are factorizable. For example, we argue in Sect. 3.9 (Observation 3.6) that clusters of highly interconnected nodes are often factorizable. Assume that a module assignment Cl is available so that each node is assigned to exactly one of Q modules. The module assignment $Cl(i) = q$ if the i th node is in the q th module. In the following, we outline when a general network adjacency A can be approximated with a parsimonious matrix $A_{CF,app}^{Cl}$ (2.15) that involves only $Q(Q - 1)/2 + n$ parameters.

Denote the module size, i.e., the number of nodes in the q th module, by $n^{(q)}$ and the $n^{(q)} \times n^{(q)}$ dimensional intramodular adjacency matrix by $A^{(q,q)}$. Assume that $A^{(q,q)}$ is approximately factorizable, i.e.,

$$A_{ij}^{(q,q)} \approx CF_i^{(q)} CF_j^{(q)}. \quad (2.12)$$

Let us now study the relationships between two modules labeled by q_1 and q_2 , respectively. In this case, the $n^{(q_1)} \times n^{(q_2)}$ matrix $A^{(q_1,q_2)}$ specifies the *inter*modular adjacencies. We find that the intermodular adjacency matrix $A_{ij}^{(q_1,q_2)}$ can sometimes be factored as follows:

$$A_{ij}^{(q_1,q_2)} \approx CF_i^{(q_1)} CF_j^{(q_2)} A_{q_1 q_2}, \quad (2.13)$$

where $A_{q_1 q_2}$ is a number, i and j denote nodes in modules q_1 and q_2 , respectively. In Sect. 6.4, we characterize *correlation* networks where (2.13) is approximately satisfied. Note that (2.12) and (2.13) can be combined as follows:

$$A_{ij}^{(q_1,q_2)} \approx CF_i^{(q_1)} CF_j^{(q_2)} A_{q_1 q_2}, \quad (2.14)$$

where $A_{q_1 q_2} = 1$ if $q_1 = q_2$. If $0 \leq A_{q_1 q_2} \leq 1$, then $A_{q_1 q_2}$ can be interpreted as an element of a $Q \times Q$ dimensional adjacency matrix $A_{modules}$ whose nodes are modules (see also Sect. 1.9).

The right-hand side of (2.14) motivates us to define an $n \times n$ dimensional matrix $A_{CF,app}^{Cl}$ whose i, j th element is given by

$$A_{CF,app,ij}^{Cl} = CF_i^{(q_1)} CF_j^{(q_2)} A_{q_1 q_2} \text{ if } i \in \text{module } q_1, j \in \text{module } q_2. \quad (2.15)$$

This leads us to the following:

Definition 2.3 (Module- and CF-based approximation of A). We refer to $A_{CF,app}^{Cl}$ (2.15) as the module- and CF-based approximation of A .

Note that $A_{CF,app}^{Cl}$ only involves $Q(Q-1)/2 + n$ parameters (comprising the upper off-diagonal elements of $A_{modules}$ and the $n = \sum_{u=1}^Q n^{qu}$ conformity parameters). If the approximation is accurate, then $A_{CF,app}^{Cl}$ provides a parsimonious parametrization of A , which is very intuitive since it is based on the underlying module structure. To quantify how well $A_{CF,app}^{Cl}$ approximates A , one can use the **module- and CF-based factorizability measure** $F(A, Cl)$ (2.19) defined below.

Let us now describe how to estimate the conformity vectors $CF^{(q_1)}, \dots, CF^{(q_Q)}$ and the intermodular adjacencies A_{q_1, q_2} . The idea is to estimate the parameters such that the resulting matrix $A_{CF,app}^{Cl}$ is as close to A as possible. Specifically, the parameters are estimated by minimizing the following objective function:

$$f(v^{(q_1)}, \dots, v^{(q_Q)}, M_{modules}) = \sum_{q_1=1}^Q \sum_{q_2=1}^Q \sum_{i \in \mathcal{M}_{q_1}} \sum_{j \in \mathcal{M}_{q_2}} \left(A_{ij}^{(q_1, q_2)} - M_{q_1, q_2} v_i^{(q_1)} v_j^{(q_2)} \right)^2, \quad (2.16)$$

where the vector $v^{(q_1)}$ is used to estimate $CF^{(q_1)}$ and the symmetric matrix $M_{modules}$ is used to estimate $A_{modules}$. We assume that the elements of $M_{modules}$ are nonnegative and that the diagonal elements equal 1. It can be convenient to reformulate the optimization problem as a maximization problem whose objective function is given by the **module-based factorizability function**:

$$F_{A,Cl}(v^{(q_1)}, \dots, v^{(q_Q)}, M_{modules}) = 1 - \frac{f(v^{(q_1)}, \dots, v^{(q_Q)}, M_{modules})}{f(0, \dots, 0)}. \quad (2.17)$$

One can easily show that if the set of parameters $v^{(q_1)}, \dots, v^{(q_Q)}, M_{modules}$ maximizes $F_{A,Cl}$, then $-v^{(q_1)}, \dots, -v^{(q_Q)}, M_{modules}$ also maximizes this function. Since the elements of $M_{modules}$ are nonnegative, the components of a maximizing conformity vector $v^{(q_1)}$ must have the same sign. Otherwise, flipping the sign of the negative components leads to a higher value of $F_{A,Cl}$. Assume that the vectors $v^{*(q_1)}, \dots, v^{*(q_Q)}$ (with nonnegative components) and the matrix $M_{modules}^*$ (with nonnegative entries) maximize $F_{A,Cl}$. If more than one set of conformity vectors

maximizes $F_{A,Cl}$ (2.17), i.e., if several solutions tie, then the maximizing set closest to the set of scaled intramodular connectivities

$$\left\{ \frac{k^{(q_1)}}{\sqrt{\text{sum}(k^{(q_1)})}}, \dots, \frac{k^{(q_Q)}}{\sqrt{\text{sum}(k^{(q_Q)})}} \right\} \quad (2.18)$$

is chosen. We define the conformity vectors $CF^{(q_1)}, \dots, CF^{(q_Q)}$ by $v^{*(q_1)}, \dots, v^{*(q_Q)}$ and the intermodular adjacency matrix $A_{modules}$ by $M_{modules}^*$. Further, we define the **module-based factorizability measure** as follows:

$$F(A, Cl) = F_{A,Cl} \left(CF^{(q_1)}, \dots, CF^{(q_Q)}, A_{modules} \right). \quad (2.19)$$

The closer $F(A, Cl)$ is to 1, the better $A_{CF,app}^{Cl}$ (2.15) approximates A . If $F(A, Cl) = 1$, then we refer to $CF^{(q_1)}, \dots, CF^{(q_Q)}, A_{modules}$ as the **module- and CF-based decomposition** of A .

An **algorithm for optimizing the two objective functions** (2.16 and 2.17) can be informed using the following comments.

First, it is natural to choose the scaled intramodular connectivity $\frac{k^{(q_1)}}{\sqrt{\text{sum}(k^{(q_1)})}}$ (2.18) as starting value for $v^{(q_1)}$.

Second, given estimates $\hat{v}^{(q_1)}$ of the conformity vectors one can derive the following explicit solution of the optimizing problem for the intermodular adjacencies

$$\hat{M}_{q_1, q_2} = \frac{\text{mean}(A^{(q_1, q_2)})}{\text{mean}(\hat{v}^{(q_1)}) \text{mean}(\hat{v}^{(q_2)})} \quad (2.20)$$

with the usual arguments from calculus (set the first derivative with respect to the conformities to zero and solve for M_{q_1, q_2}). This suggests to optimize the objective function by iterating between two steps: one step estimates $v^{(q_1)}$ and the other estimates M_{q_1, q_2} (using (2.20)).

Third, an *approximate* solution to the optimization problem can be found by ignoring intermodular adjacencies when it comes to estimating $v^{(q_1)}$. In this case, the algorithm for finding conformity vectors (in Sect. 2.2.1) can be used to find an approximate estimate $\hat{v}^{(q_1)} = \hat{C}F^{(q_1)}$ based on $A^{(q_1, q_1)}$ alone. An algorithm for finding these approximate solutions is implemented in the R function `conformityDecomposition`. It is worth repeating that these approximate estimates are typically different from the true solutions of the optimization problem. **John Ranola** and **Kenneth Lange** have developed a minorization–majorization (MM) algorithm for finding a solution of the optimization problem, which is implemented in the R function `propensityDecomposition` (Ranola et al. 2011). We briefly mention that one can extend the optimization problem using an objective function that includes the module assignment Cl as an additional parameter. Optimizing this extended objective function with regard to Cl , the conformities, and the intermodular

adjacencies leads to a CF based clustering algorithm, which is implemented in the function `propensityClustering`. Example R code illustrating the use of `propensityDecomposition` and `propensityClustering` can be found in the help files of these functions.

2.4 Exercises

- Exercise regarding the uniqueness of conformity in case of an exactly factorizable network. Specifically, prove the following statement. If A is an $n \times n$ ($n \geq 3$) dimensional adjacency matrix with positive entries ($A_{ij} > 0$), then the system of (2.1)

$$A_{ij} = CF_i CF_j \text{ for all } i \neq j$$

has at most one solution CF with positive entries. If the solution exists, it is given by

$$CF_i = \left(\frac{p_i}{(\prod_{h=1}^n p_h)^{1/(2(n-1))}} \right)^{\frac{1}{n-2}}, \quad (2.21)$$

where $p_i = \prod_{j=1}^n A_{ij}$ denotes the **product connectivity** of the i th node. Hint: By assumption, we have $A_{ij} = CF_i CF_j$ for a positive vector CF and $n \geq 3$. Multiplying both sides of (2.1) yields $\prod_h \prod_{l \neq h} A_{lh} = \prod_h \prod_{l \neq h} CF_l CF_h = (\prod_{l=1}^n CF_l)^{2(n-1)}$. Since $\prod_{l=1}^n CF_l$ is positive, we find

$$\prod_{l=1}^n CF_l = \left(\prod_h \prod_{l \neq h} A_{lh} \right)^{\frac{1}{2(n-1)}}.$$

Similarly, eliminating the i th row and column from A yields

$$\prod_{l \neq i} CF_l = \left(\prod_{h \neq i} \prod_{l \neq h, i} A_{lh} \right)^{\frac{1}{2(n-2)}} = \left(\prod_h \prod_{l \neq h} A_{lh} / \left(\prod_{l \neq i} A_{li} \right)^2 \right)^{\frac{1}{2(n-1)}}.$$

Since $CF_i = \prod_{l=1}^n CF_l / \prod_{l \neq i} CF_l$, we conclude that CF_i is uniquely defined by

$$CF_i = \frac{(\prod_h \prod_{l \neq h} A_{lh})^{\frac{1}{2(n-1)}}}{(\prod_{h \neq i} \prod_{l \neq h, i} A_{lh})^{\frac{1}{2(n-2)}}} = \left(\frac{p_i}{(\prod_{h=1}^n p_h)^{1/(2(n-1))}} \right)^{\frac{1}{n-2}}.$$

- Exercise regarding the module- and CF-based approximation (decomposition) of a block diagonal network A . Consider an unweighted block diagonal adjacency matrix with two blocks. The first and second blocks contain $n^{(1)}$ and $n^{(2)}$ nodes, respectively. Each nonzero element of the first and second blocks equals b_1 and b_2 , respectively.

- (i) Calculate the module- and CF-based approximation of A if CI indicates block (module) membership. Hint: See Sects. 3.11 and 2.3. Determine $CF^{(1)}$, $CF^{(2)}$, and the 2×2 dimensional matrix $A_{modules}$.
- (ii) Is the result of (i) a decomposition of A ? Hint: Determine whether the module-based factorizability measure $F(A, CI)$ equals 1.
- (iii) Determine the module- and CF-based decomposition of a general block diagonal matrix with Q blocks of sizes $n^{(1)}, \dots, n^{(Q)}$.

References

- Deeds EJ, Ashenberg O, Shakhnovich EI (2006) A simple physical model for scaling in protein–protein interaction networks. *Proc Natl Acad Sci USA* 103(2):311–316
- Dong J, Horvath S (2007) Understanding network concepts in modules. *BMC Syst Biol* 1(1):24
- Gifi A (1990) Nonlinear multivariate analysis. Wiley, Chichester, England
- Horn RA, Johnson CR (1991) Topics in matrix analysis. Cambridge University Press, Cambridge
- Lange K (2004) Optimization. Springer, New York
- de Leeuw J, Michailidis G (2000) Block relaxation algorithms in statistics. *J Comput Graph Stat* 9:26–31
- Ranola JMO, Ahn S, Sehl ME, Smith DJ, Lange K (2010) A Poisson model for random multi-graphs. *Bioinformatics* 26(16):2004–2011
- Ranola JMO, Langfelder P, Song L, Horvath S, Lange K (2011) An MM algorithm for module and propensity based decomposition of a network. UCLA Technical Report
- Servedio VDP, Caldarelli G, Butta P (2004) Vertex intrinsic fitness: How to produce arbitrary scale-free networks. *Phys Rev E – Stat Nonlin Soft Matter Phys* 70(5):056126



<http://www.springer.com/978-1-4419-8818-8>

Weighted Network Analysis
Applications in Genomics and Systems Biology
Horvath, S.
2011, XXIII, 421 p., Hardcover
ISBN: 978-1-4419-8818-8