universität wien

MASTERARBEIT | MASTER'S THESIS

Titel | Title

Benchmarking edge-specific algorithms for metabolic and microbial graphs

verfasst von | submitted by

Dipl.-Ing. Leopold Zehetner BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 863
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Biologische Chemie
Betreut von | Supervisor:	Univ.-Prof. Dipl.-Ing. Dr. Jürgen Zanghellini

Acknowledgments

Above all, I want to express my deep gratitude to Univ.-Prof. Dipl.-Ing. Dr. Jürgen Zanghellini, Professor of Biochemical Network Analysis, for his inspirational way of teaching, which encouraged me to perform my master's project in his research group. His enthusiastic manner motivated me over the whole time of my master's project and will have a major impact on my future career. Besides the kind support, which was a cornerstone during this project, I am exceptionally thankful for his advice and feedback on my thesis.

Additionally, I would like to sincerely thank my co-supervisor, Mgr. Diana Szeliova, PhD, for the inspiring meetings, through which I could gain new knowledge, and her availability, whenever I needed support.

Besides, I would also like to thank all my colleagues for the great atmosphere and the interesting presentations.

Finally, I want to thank my whole family and friends for their daily support and their belief in my work.

Abstract

Biological networks present unique challenges due to their inherent properties of directed edges and weighted nodes and edges. Effective algorithms must consider these complexities to yield meaningful insights. Traditional algorithms often either highlight cofactor-associated hubs or produce inconsistent results. To address these issues, we applied the Forman-Ricci curvature (FRC) to both metabolic and microbial graphs and compared its performance with other algorithms such as edge betweenness centrality, edge pagerank, edge eigenvector centrality, edge cluster coefficient, edge local assortativity, and Ollivier-Ricci curvature (ORC). Additionally, we enhanced the FRC algorithm by normalizing edge numbers based on node connectivity, aiming to provide a more detailed and balanced network analysis. We first evaluated the algorithms on small and medium-sized random networks, where ORC demonstrated superior performance in maintaining the weighted largest connected compound (LCC) after sequential edge removal. FRC's performance improved significantly with edge number normalization, closely matching ORC, while edge betweenness centrality performed the worst. Runtime analysis indicated that FRC and ORC were significantly slower, which may be a drawback for larger networks. Applying these algorithms to a genome-scale metabolic model (GSMM) of *E. coli* under glucose-rich and fucose-rich conditions, ORC again maintained the LCC size most effectively, followed by edge pagerank and edge cluster coefficient. Principal component analysis (PCA) of edge rankings showed consistent clustering patterns for FRC and ORC. Edge number normalization enhanced FRC's performance, with limited impact on other algorithms. We also examined edge-specific value distributions, finding that FRC displayed significant deviations, particularly in the fucose network, indicating reduced stability. Statistical measures confirmed that FRC outperformed other algorithms in distinguishing between glucose and fucose conditions when distributions were normalized. In microbial networks from Levy et al. [1], consisting of 154 microbes and 23,188 interactions, FRC showed robust performance, further improved by edge number normalization. FRC effectively differentiated between various states (healthy lean, healthy obese, inflammatory bowel disease (IBD) lean, and IBD obese), with the microbial network in IBD obese patients being the least stable. In conclusion, our benchmarking study demonstrated that FRC, especially when augmented with edge number normalization, is a powerful tool for analyzing metabolic and microbial graphs. It consistently outperformed other algorithms in maintaining network integrity and distinguishing between different network conditions. Future studies should focus on larger datasets and further validation of these findings in more complex biological networks.

Abstract (deutsch)

Biologische Netzwerke stellen aufgrund ihrer inhärenten Eigenschaften von gerichteten Kanten und gewichteten Knoten und Kanten einzigartige Herausforderungen dar. Effektive Algorithmen müssen diese Komplexitäten berücksichtigen, um aussagekräftige Einblicke zu liefern. Traditionelle Algorithmen heben oft entweder Cofaktor-assoziierte Hubs hervor oder liefern inkonsistente Ergebnisse. Um diese Probleme zu adressieren, haben wir die Forman-Ricci-Krümmung (FRC) sowohl auf metabolische als auch auf mikrobielle Netzwerke angewendet und ihre Leistung mit anderen Algorithmen wie Kanten-Betweenness-Zentralität, Kanten-Pagerank, Kanten-Eigenvektor-Zentralität, Kanten-Cluster-Koeffizient, lokaler Kanten-Assortativität und Ollivier-Ricci-Krümmung (ORC) verglichen. Zusätzlich haben wir den FRC-Algorithmus durch die Normalisierung der Kantenzahl basierend auf der Knotenverbindung verbessert, um eine detailliertere und ausgewogenere Netzwerkanalyse zu ermöglichen.

Wir haben die Algorithmen zunächst an kleinen und mittelgroßen zufällig generierten Netzwerken bewertet, wobei ORC eine überlegene Leistung bei der Aufrechterhaltung der gewichteten LCC nach sequentieller Kantenentfernung zeigte. Die Leistung von FRC verbesserte sich signifikant mit der Kantenzahl-Normalisierung und kam der von ORC nahe, während die Kanten-Betweenness-Zentralität am schlechtesten abschnitt. Laufzeitanalysen zeigten, dass FRC und ORC erheblich langsamer waren, was bei größeren Netzwerken nachteilig sein könnte.

Bei der Anwendung dieser Algorithmen auf ein GSMM von *E. coli* unter glucose- und fucose-reichen Bedingungen zeigte ORC erneut die effektivste Aufrechterhaltung der LCC-Größe, gefolgt von Kanten-Pagerank und Kanten-Cluster-Koeffizient. PCA der Kantenrankings zeigte konsistente Cluster-Muster für FRC und ORC. Die Kantenzahl-Normalisierung verbesserte die Leistung von FRC, hatte jedoch begrenzte Auswirkungen auf andere Algorithmen. Wir untersuchten auch die verteilten kanten-spezifischen Werte und stellten fest, dass FRC signifikante Abweichungen zeigte, insbesondere im Fucose-Netzwerk, was auf eine reduzierte Stabilität hindeutet. Statistische Maßnahmen bestätigten, dass FRC andere Algorithmen übertraf, wenn es darum ging, zwischen glucose- und fucose-reichen Bedingungen zu unterscheiden, wenn die Verteilungen normalisiert wurden.

In mikrobiellen Netzwerken von Levy et al. [1], die aus 154 Mikroben und 23.188 Interaktionen bestanden, zeigte FRC eine robuste Leistung, die durch die Kantenzahl-Normalisierung weiter verbessert wurde. FRC unterschied effektiv zwischen verschiedenen Zuständen (gesund schlank, gesund fettleibig, IBD schlank und IBD fettleibig), wobei das mikrobielle Netzwerk bei IBD fettleibigen Patienten am instabilsten war.

Abschließend zeigte unsere Benchmark-Studie, dass FRC, insbesondere in Kombination mit der Kantenzahl-Normalisierung, ein leistungsfähiges Werkzeug zur Analyse von metabolischen und mikrobiellen Netzwerken ist. Es übertraf konsistent andere Algorithmen in der Aufrechterhaltung der Netzwerkintegrität und der Unterscheidung zwischen verschiedenen Netzwerkbedingungen. Zukünftige Studien sollten sich auf größere Datensätze und die weitere Validierung dieser Ergebnisse in komplexeren biologischen Netzwerken konzentrieren.

Introduction

Networks are fundamental constructs used to represent relationships between entities. A network is composed of nodes (or vertices) which denote the entities, and edges (or links) which signify the interactions or connections between these entities [2]. Networks can be categorized based on the nature of their edges: directed networks have edges with a specific direction, indicating a one-way relationship between nodes, while undirected networks have edges that indicate a bidirectional relationship, implying mutual interaction [2,3]. Additionally, networks can be weighted or unweighted. In weighted networks, edges carry a numerical value that represents the strength or capacity of the connection, whereas in unweighted networks, all edges are considered equal, devoid of any additional quantitative information [3,4]. Networks provide a framework for understanding complex relationships, with prominent examples including social networks, which map social relationships between individuals or organizations; email networks, where individuals are nodes and email exchanges represent edges; or various biological networks [3]. Biological networks, in particular, have become a vital area of study, encompassing several types, such as protein-protein interaction (PPI) [5–7], gene regulatory networks [8], genome-scale metabolic models (GSMMs) [9,10], signal transduction networks [11], or ecological networks [1,12,13].

In the last two decades, GSMMs have been introduced to analyze intracellular processes [14]. GSMMs, which mathematically represent a cell's metabolism, have been instrumental since their first development for *H. influenzae* [15]. They have been used to optimize bioprocesses [16–19], studying diseases [20–22], or microbial commmunities [23,24]. Protocols for reconstructing GSMMs, aided by automatic pipelines, have resulted in hundreds of models, including highly curated global GSMMs for model organisms like H. sapiens [25,26], C. griseus [27], M. musculus [28], and others [28]. These models allow cell- or tissue-specific reconstructions based on transcriptomes and metabolic exchange rates [17,29,30]. These models can be used to optimize certain objective functions, such as biomass for growth [17,29] or production of certain molecules [31,32]. Several optimization algorithms have been developed, including flux balance analysis (FBA) [33], parsimonious flux balance analysis (PFBA) [34], flux variability analysis (FVA) [35], or resource allocation analysis (RBA) [36,37]. FBA is a mathematical approach used to predict the flow of metabolites through a metabolic network under steady-state conditions by optimizing an objective function, typically representing cellular growth or biomass production, subject to constraints such as mass balance [33]. PFBA extends FBA by minimizing the total flux through the network, selecting the most efficient flux distribution that reflects a parsimonious use of cellular resources, in addition to optimizing a primary objective function [34]. FVA determines the range of possible flux values for each reaction in the network while satisfying an optimal solution to an objective function, involving the assessment of minimum and maximum fluxes under given constraints [35]. Most recently developed, RBA simulates the allocation of cellular resources, such as ribosomes, enzymes, and energy, in a metabolic network, incorporating physiological limitations and optimizing an objective function to ensure efficient resource use within cellular capacity [36,37]. In contrast to optimization approaches, flux sampling is a computational technique that explores the range of possible flux distributions within a metabolic network given constraints, providing an overview of all feasible metabolic states without the need for an objective function [38]. Despite all of these approaches have been benchmarked and applied in various ways, it remains difficult to analyze GSMMs as metabolic graphs [39], which allows to study properties such as degree distribution, centrality, clustering coefficients, and shortest paths. These properties can reveal important aspects of the network's organization and functionality. By analyzing centrality measures (e.g., betweenness centrality, degree centrality), it is possible to identify key metabolites and reactions that are crucial for the network's integrity and function [9,40,41]. The main challenge in converting GSMMs into metabolic graphs is that there is no standardized procedure of how to convert a GSMM into a metabolic graph [42,43]. One approach involves using metabolites as nodes and reactions as edges, enabling the identification of key metabolites within the metabolic networks [10,44]. This method faces challenges since the most critical metabolites (hubs) are often cofactors like ATP, ADP, NAD, NADH, or other small molecules such as H2O, H+, or O2, which are involved in many reactions [45]. To avoid these

molecules as hubs, they can be excluded from the analysis, but this significantly alters the metabolic network [40,41]. Another approach is to interpret reactions as nodes and metabolites as connecting edges [9,46,47]. This method has the advantage that node deletions can be linked to specific genes associated with the corresponding reactions. However, this approach results in larger graphs due to the higher number of reactions compared to metabolites, increasing the complexity of the analysis [9]. The lack of standardization in converting GSMMs to metabolic graphs is a significant pitfall. To address the challenges of metabolic network generation and the question of whether nodes should represent metabolites or reactions, it is necessary to apply algorithms that allow the analysis of node and edge weights in directed networks simultaneously.

Aside from metabolic graphs, network analysis in microbial communities is a powerful tool for understanding the complex interactions and dynamics within these ecosystems [12]. Co-occurrence networks, which are constructed based on the presence or absence of species across samples, are a common approach to infer potential interactions among microbial taxa [1]. These networks are typically reconstructed using various statistical methods, such as correlation-based approaches, graphical lasso, and association rule mining, to identify significant associations between species [12,48]. Once constructed, these networks can be analyzed to identify key structural properties, such as hubs, modules, and network motifs, which can provide insights into the roles of specific taxa and the overall stability and functionality of the microbial community [49,50]. However, there are several limitations to these algorithms. One major challenge is the compositional nature of microbiome data, where relative abundances can introduce biases, making it difficult to infer true ecological interactions [51]. Additionally, the heterogeneity and high dimensionality of microbial datasets can complicate the interpretation of network structures and the identification of meaningful biological patterns [52]. Furthermore, co-occurrence does not necessarily imply direct ecological interactions, and significant associations observed in networks may be driven by indirect effects or shared environmental factors rather than true biological interactions [53]. Despite these challenges, network analysis remains a valuable approach for exploring the intricate web of interactions within microbial communities. Levy, et al. [1] integrated metagenomic-based compositional data with GSMMs to predict levels of metabolic competition and metabolic complementarity among microbiome species, using *in silico* metabolic network models to compare predicted interaction measures to species co-occurrence. This approach revealed that species tend to co-occur more frequently with those they strongly compete with, suggesting habitat filtering as a dominant assembly force in the gut microbiome. Guo, et al. [54] utilized co-occurrence network analysis to study microbial interactions in anaerobic digestion systems. The study constructed separate networks for different operational conditions and found that network topological properties, such as clustering coefficient and average path length, were correlated with reactor performance metrics like hydrolysis and methanogenesis efficiency. The study highlighted the importance of low-abundance genera, which, despite their low relative abundance, played central roles in maintaining network stability and functionality. Estrada-Pena, et al. [55] applied network analysis tools to disentangle the microbial communities of ticks and their hosts. The study constructed networks based on the co-occurrence of bacterial genera and found that the biotope significantly influenced the bacterial communities. The analysis revealed that certain bacterial genera acted as central nodes, playing crucial roles in the network's structure and stability. The study also demonstrated that the microbiome of ticks and their hosts were only slightly related, emphasizing the importance of environmental factors in shaping microbial communities. Zamkovaya and colleagues [56] investigated the ecological importance of "microbial dark matter" by creating microbial co-occurrence networks from 16S rRNA gene sequencing data obtained from extreme aquatic environments. They compared networks that included and excluded unknown taxa, discovering that these unknown taxa frequently served as central hubs, highlighting their essential roles in ecosystem structure. To quantitatively evaluate the significance of these unknown taxa, the study employed network centrality metrics such as degree, betweenness, and closeness. Rio-Hortega, et al. [57] employed network analysis to compare plant community assembly between native and invaded ranges. The study constructed co-occurrence networks to identify key species and their roles within the community. The analysis revealed that

certain species in the invaded range acted as hubs, displaying higher connectivity and playing crucial roles in maintaining network structure. The study emphasized the utility of network analysis in understanding community assembly and the impact of biological invasions. In pollination networks, the study by Gonzalez, et al. [58] examined the relationship between species generalization and centrality indices such as closeness and betweenness centrality. The study found that generalist species often had high centrality scores, indicating their importance in maintaining network cohesiveness and stability. The analysis demonstrated that these species acted as connectors, linking otherwise unconnected subnetworks and playing key roles in the overall network structure. These studies collectively demonstrate the power of network analysis in elucidating the complex interactions within microbial and ecological communities. By applying various centrality measures and constructing detailed co-occurrence networks, it is possible to identify key species and understand their roles in maintaining ecosystem stability and functionality. This approach provides a robust framework for studying community dynamics, predicting responses to disturbances, and informing conservation and management strategies. However, all these studies rely only on node-centric measures, where relative species abundances served as node weights, neglecting the metabolic interaction of species. This limitation might have been occured due to the lack of algorithms, analyzing weighted edges, and nodes simultaneously.

Recent advancements have introduced edge-based metrics, particularly the FRC, which offers a novel perspective on network structure and dynamics [59,60]. FRC [61] quantifies the relational character of nodes through the analysis of edges rather than nodes, providing a more nuanced understanding of network topology [59]. FRC has been successfully applied to both undirected and directed networks, revealing significant insights into their structural properties. For instance, in directed networks, FRC has been shown to incorporate node weights, edge weights, and edge direction, thereby enabling the characterization of complex directed networks [62,63]. FRC has been employed to detect redundant and bottleneck reactions in hypergraphs, highlighting its potential for identifying critical network components [59]. In social networks, such as the Facebook network, FRC has been used to study the topology of social interactions, revealing the importance of directionality in information flow [60]. Comparative studies have also been conducted to evaluate the efficacy of FRC against other discrete notions of Ricci curvature, such as Ollivier-Ricci curvature [64]. These studies have shown that while both measures are highly correlated in many networks, FRC offers computational advantages, making it more suitable for large-scale network analysis [65,66]. Furthermore, augmentations of FRC have been proposed to enhance its applicability in community detection, providing a computationally cheaper alternative to Ollivier-Ricci curvature-based methods [67]. FRC has been utilized to quantify cellular pluripotency and pathway robustness. Research has shown a direct relationship between Ricci curvature and network entropy, implying that curvature can be used as a marker of cellular pluripotency [68]. Additionally, local curvature analysis has identified critical genes and pathways in cancer networks, underscoring the potential of FRC in biomedical research [68]. FRC is an adaptation of the classical Ricci curvature for discrete settings, particularly for networks or graphs [61]. For a given edge $e = (u, v)$ in a graph, the FRC $F(e)$ is defined as:

$$F(e) = \omega(e)\left(\frac{1}{\omega(u)} + \frac{1}{\omega(v)}\right) - \sum_{e_i \sim u} \frac{\omega(e)}{\sqrt{\omega(e)\omega(e_i)}} - \sum_{e_j \sim v} \frac{\omega(e)}{\sqrt{\omega(e)\omega(e_j)}} \tag{1}$$

where:

- $\omega(e)$ is the weight of the edge e,
- $\omega(u)$ and $\omega(v)$ are the weights of the nodes u and v,
- $e_i \sim u$ indicates edges e_i incident to the node u,
- $e_j \sim v$ indicates edges e_j incident to the node v.

Ollivier-Ricci curvature is another adaptation of the classical Ricci curvature for graphs, based on optimal transport theory [64]. This curvature considers the transportation cost of moving probability measures between neighboring nodes. For two adjacent nodes u and v in a graph, the Ollivier-Ricci curvature $\kappa(u, v)$ is defined as:

$$\kappa(u, v) = 1 - \frac{W_1(\mu_u, \mu_v)}{d(u, v)} \tag{2}$$

where:

- W_1 is the Wasserstein distance (earth mover's distance) between the probability measures μ_u and μ_v,
- μ_u and μ_v are the probability measures supported on the neighbors of u and v, respectively,
- $d(u, v)$ is the graph distance between nodes u and v.

In this study, we implemented FRC to analyze directed networks, derived from GSMMs or microbiome data, with weighted edges and nodes. Initially, we constructed random networks to compare FRC [61,65] against established network analysis algorithms, including edge betweenness centrality, edge pagerank, edge eigenvector centrality, edge cluster coefficient, edge local assortativity, and ORC [64]. Each algorithm's effectiveness was evaluated by systematically removing edges starting from the top-ranked edges for every method and calculating the size of the largest connected compound (LCC) [69]. Subsequently, we applied these algorithms to biological networks, including transcriptome-weighted metabolic networks from E. coli treated with antibiotics [70], to determine if the algorithmic rankings could conserve a similar clustering pattern in principal component analysis (PCA) [71] compared to the entire transcriptome. Furthermore, we examined networks with transcriptome weights for edges and metabolite weights for nodes, measured under fucose- and glucose-rich media [72]. In the context of microbial communities, we generated a network with metabolic complementarity scores as edge weights between species-assigned nodes, integrating node weights from co-occurrence values [1]. Additionally, we adapted the FRC [61] algorithm to account for hub nodes by normalizing by the number of edges per node, enhancing the algorithm's performance and making it more comparable to the best-performing Ollivier-Ricci curvature [64]. Our findings demonstrate that FRC is highly effective for analyzing metabolic and microbial networks with weighted edges and nodes, outperforming other commonly used algorithms. By normalizing for the number of edges each node is connected to, we improved the performance of FRC, particularly in the selected networks, compared to the original algorithm.

Methods

Edge-specific Network Algorithms

Edge Betweenness Centrality

Edge betweenness centrality [73] measures the importance of an edge based on the number of shortest paths that pass through it. Specifically, the edge betweenness centrality of an edge e is defined as the sum of the fraction of all-pairs shortest paths that pass through e. The edge betweenness centrality $C_B(e)$ for an edge e is given by:

$$C_B(e) = \sum_{u \neq v \in V} \frac{\sigma_{uv}(e)}{\sigma_{uv}} \tag{3}$$

where:

- V is the set of nodes in the network.
- σ_{st} is the total number of shortest paths from node u to node v.
- $\sigma_{st}(e)$ is the number of those shortest paths that pass through edge e.

In the context of weighted networks, the shortest paths are determined by the sum of the weights along the paths, and not just the number of edges.

$$C_B(e) = \sum_{u \neq v \in V} \frac{\delta_{uv}(e)}{\delta_{uv}} \tag{4}$$

where:

- δ_{uv} is the sum of the weights of the shortest paths from node u to node v.
- $\delta_{uv}(e)$ is the sum of the weights of the shortest paths from u to v that pass through edge e.

Hence, an edge with a high edge betweenness centrality indicates that it acts as a critical bridge in the network, facilitating communication between various parts of the network. This can highlight potential points of failure or bottlenecks in the network structure. The computation of edge betweenness centrality involves finding the shortest paths between all pairs of nodes and then counting how many of these paths traverse each edge.

Edge Pagerank

Pagerank is a widely used algorithm for measuring the importance of nodes in a network. Originally developed by Google to rank web pages [74], pagerank assigns a numerical value to each node based on its connections. The fundamental idea is that a node is important if it is linked to by other important nodes.

The general pagerank formula is:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in M(v)} \frac{PR(u)}{L(u)} \tag{5}$$

where:

- $PR(v)$ is the pagerank of node v.
- d is the damping factor, set to 0.85.
- N is the total number of nodes.

- $M(v)$ is the set of nodes that link to v.
- $L(u)$ is the number of outbound links from node u.

For directed networks with weighted edges, the pagerank algorithm must be adapted to account for edge weights. This involves distributing the pagerank proportionally based on the weights of the edges. The adapted pagerank formula for a node v in a weighted, directed network is:

$$PR(v) = \frac{1-d}{N} + d \sum_{u \in M(v)} \frac{w_{uv} \cdot PR(u)}{W(u)} \tag{6}$$

where:

- w_{uv} is the weight of the edge from node u to node v.
- $W(u)$ is the sum of the weights of all outbound edges from node u.

To apply the pagerank algorithm to edges, we first computed the pagerank for each node and then distribute the pagerank to edges proportionally based on their weights.

Edge Eigenvector Centrality

Edge eigenvector centrality measures the importance of an edge based on the eigenvector centralities of the nodes [75] it connects. Specifically, the edge eigenvector centrality of an edge e = (u, v) is defined as the average of the eigenvector centralities of the nodes u and v. The edge eigenvector centrality $C_E(e)$ for an edge e is given by:

$$C_E(e) = \frac{C_E(u) + C_E(v)}{2} \tag{7}$$

where:

- $C_E(u)$ is the eigenvector centrality of node u.
- $C_E(v)$ is the eigenvector centrality of node v.

The eigenvector centrality of a node measures its influence in the network by considering not only the number of its connections but also the centrality of its neighbors. In mathematical terms, the eigenvector centrality C_E of a node u is given by the principal eigenvector of the adjacency matrix A of the graph G:

$$A\mathbf{C_E} = \lambda \mathbf{C_E} \tag{8}$$

where:

- A is the adjacency matrix of the graph G.
- $\mathbf{C_E}$ is the vector of eigenvector centralities for all nodes in the network.
- λ is the largest eigenvalue of the adjacency matrix A.

In the context of weighted networks, the weights of the edges are considered in the adjacency matrix, affecting the calculation of the eigenvector centralities of the nodes. The edge eigenvector centrality provides insights into the significance of an edge by incorporating the influence of its endpoints in the network structure. It highlights edges that connect highly influential nodes, which can be critical for understanding the flow of information or connectivity within the network.

Edge Clustering Coefficient

Edge clustering coefficient measures the clustering tendency of an edge based on the clustering coefficients [76] of the nodes it connects. Specifically, the edge clustering coefficient of an edge $e = (u, v)$ is defined as the average of the clustering coefficients of the nodes u and v. The edge clustering coefficient $C_C(e)$ for an edge e is given by:

$$C_C(e) = \frac{C_C(u) + C_C(v)}{2} \tag{9}$$

where:

- $C_C(u)$ is the clustering coefficient of node u.
- $C_C(v)$ is the clustering coefficient of node v.

The clustering coefficient of a node measures the extent to which its neighbors form a complete graph (i.e., how many of its neighbors are also neighbors with each other). In mathematical terms, the clustering coefficient C_C of a node u is given by:

$$C_C(u) = \frac{2T(u)}{k_u(k_u - 1)} \tag{10}$$

where:

- $T(u)$ is the number of triangles through node u.
- k_u is the degree of node u, i.e., the number of edges connected to u.

In the context of weighted networks, the weights of the edges are considered in the calculation of the clustering coefficients of the nodes. The edge clustering coefficient provides insights into the local cohesiveness of the network by incorporating the clustering tendencies of its endpoints. It highlights edges that connect nodes with a high tendency to form local clusters, which can be crucial for understanding the local density and robustness of the network structure.

Local Assortiativity

Local assortativity [77] measures the tendency of edges to connect nodes with similar properties, such as degree or weight. Specifically, local degree assortativity and local weight assortativity are considered. The combined local assortativity of an edge $e = (u, v)$ is computed by summing the degree and weight assortativity measures.

First, the local degree assortativity $A_D(e)$ for an edge e is given by:

$$A_D(e) = (d_u - d_v)^2 \tag{11}$$

where:

- d_u is the degree of node u.
- d_v is the degree of node v.

Next, the local weight assortativity $A_W(e)$ for an edge e is given by:

$$A_W(e) = (w_{uv} - \bar{w}_u)^2 + (w_{uv} - \bar{w}_v)^2 \tag{12}$$

where:

- w_{uv} is the weight of the edge e.
- $\bar{w}_u$ is the average weight of edges connected to node u.

- $\bar{w}_v$ is the average weight of edges connected to node v.

The average weight of edges connected to a node u is computed as:

$$\bar{w}_u = \frac{1}{|\{n : (u,n) \in E\}|} \sum_{n:(u,n)\in E} w_{un} \tag{13}$$

where:

- $\{n : (u,n) \in E\}$ is the set of neighbors of node u.
- w_{un} is the weight of the edge between nodes u and n.

The combined local assortativity $A(e)$ for an edge e is given by:

$$A(e) = A_D(e) + A_W(e) \tag{14}$$

The local assortativity measures provide insights into the homophily of the network, where high values indicate that edges tend to connect nodes with similar degrees or weights. This can be useful for understanding the local structure and organization of the network, as well as identifying patterns of connectivity based on node properties.

Ollivier-Ricci Curvature (ORC)

ORC [64] measures the curvature of an edge in a graph based on optimal transport theory, which can provide insights into the local geometric properties of the network. Specifically, the ORC of an edge $e = (u, v)$ is computed using the Wasserstein distance between the probability distributions of the neighbors of the nodes u and v.

The ORC $\kappa(e)$ for an edge e is given by:

$$\kappa(e) = 1 - \frac{W(u,v)}{d(u,v)} \tag{15}$$

where:

- $W(u,v)$ is the Wasserstein distance between the probability distributions of the neighbors of nodes u and v.
- $d(u,v)$ is the weight of the edge e.

To compute the Wasserstein distance $W(u,v)$, we define the probability distributions of the neighbors of nodes u and v:

$$p_u = \frac{[w_{un_1}, w_{un_2}, \ldots, w_u]}{\sum_{n\in N(u)} w_{un} + w_u} \tag{16}$$

$$p_v = \frac{[w_{vm_1}, w_{vm_2}, \ldots, w_v]}{\sum_{m\in N(v)} w_{vm} + w_v} \tag{17}$$

where:

- $N(u)$ is the set of neighbors of node u.
- w_{un} is the weight of the edge between node u and its neighbor n.
- w_u is the weight of node u (default is 1.0 if not specified).

The cost matrix C for transporting probability mass between the neighbors of u and v is constructed as:

$$C_{ij} = \begin{cases} w_{n_i m_j}, & \text{if } (n_i, m_j) \in E \\ 10.0, & \text{otherwise} \end{cases} \tag{18}$$

where E is the set of edges in the graph. The optimal transport plan is determined using linear programming, and the Wasserstein distance $W(u, v)$ is the cost of this optimal transport plan.

The ORC captures the notion of curvature in a discrete setting, where positive curvature indicates a locally cohesive structure, and negative curvature indicates a locally expansive structure. This measure can be particularly useful for identifying bottlenecks or bridges in the network, as well as for understanding the overall geometric properties of the graph.

Forman-Ricci Curvature (FRC)

FRC [61] measures the curvature of an edge in a graph, which can provide insights into the local geometric properties of the network. Specifically, the FRC of an edge $e = (u, v)$ is computed using the weights of the edge and its incident nodes, as well as the weights of the edges connected to these nodes. The FRC $F(e)$ for an edge e is given by:

$$F(e) = w_e \left(\frac{w_u}{w_e} + \frac{w_v}{w_e} - \sum_{(u',u) \in E_u} \frac{w_u}{\sqrt{w_e \cdot w_{(u',u)}}} - \sum_{(v,v') \in E_v} \frac{w_v}{\sqrt{w_e \cdot w_{(v,v')}}} \right) \tag{19}$$

where:

- w_e is the weight of edge e.
- w_u is the weight of the node u.
- w_v is the weight of the node v.
- E_u is the set of edges incident to node u excluding edge e.
- E_v is the set of edges incident to node v excluding edge e.
- $w_{(u',u)}$ is the weight of an edge in E_u.
- $w_{(v,v')}$ is the weight of an edge in E_v.

The first two terms inside the parentheses represent the contributions from the weights of the incident nodes normalized by the weight of the edge e. The sum terms represent the contributions from the edges connected to the incident nodes, scaled by the geometric mean of their weights and the weight of e.

The FRC captures the notion of curvature in a discrete setting, where positive curvature indicates a locally cohesive structure, and negative curvature indicates a locally expansive structure. This measure can be particularly useful for identifying bottlenecks or bridges in the network, as well as for understanding the overall geometric properties of the graph.

Normalization for edge number

In an additional attempt to refine our network analysis, we augmented the algorithms by accounting for the number of edges connected to the nodes of each edge. For a given edge from node u to node v, we counted all incoming edges to node u and all outgoing edges from node v. The sum of these counts provided the normalization factor, by which the final edge-specific value was divided. This adjustment ensured that the resulting value for each edge from each algorithm was considered relative to the connectivity of the nodes it linked, thereby providing a more balanced analysis of edge

importance. This is of special importance in biological networks, since certain cofactors or small molecules, such as NADH, NAD, ATP, etc. participate in many reactions, making them more important due to the sheer amount of connected edges [40, 41, 45]. The only exception to this general approach was made for the FRC, where the original formula was adapted as follows:

$$F(e) = w_e \left(\frac{w_u}{w_e} + \frac{w_v}{w_e} - \frac{\sum_{(u',u)\in E_u} \frac{w_u}{\sqrt{w_e \cdot w_{(u',u)}}}}{\#E_u^{in}} - \frac{\sum_{(v,v')\in E_v} \frac{w_v}{\sqrt{w_e \cdot w_{(v,v')}}}}{\#E_v^{out}} \right) \tag{20}$$

where:

- w_e is the weight of edge e.
- w_u is the weight of the node u.
- w_v is the weight of the node v.
- E_u is the set of edges incident to node u excluding edge e.
- E_v is the set of edges incident to node v excluding edge e.
- $\#E_u^{in}$ is the number of ingoing edges to node u.
- $\#E_v^{out}$ is the number of outgoing edges from node v.
- $w_{(u',u)}$ is the weight of an edge in E_u.
- $w_{(v,v')}$ is the weight of an edge in E_v.

Comparison of Algorithms

To compare the performance and characteristics of different network analysis algorithms, we first obtained a value for each edge in the network from each algorithm. These values were then used to assign a rank to every edge for each algorithm, allowing for comparability across methods while preserving the order determined by each algorithm. The rankings obtained were subsequently used to perform a PCA [71], facilitating the identification of similarities and differences among the algorithms. We utilized the `scikit-learn` [78] package in Python. The PCA provided a visual representation of the relationships between the different algorithms based on their edge rankings, highlighting clusters and patterns indicative of algorithmic similarities and differences.

To further evaluate the effectiveness of each algorithm, we employed an adapted approach to analyze the weighted LCC [69] by iteratively removing the highest-ranked edge. This method aimed to identify which algorithm was most successful in pinpointing the most crucial edges in the network. The process involved several key steps: First, we defined a method to calculate the weighted degree of a node. This method sums the weights of all incoming and outgoing edges for a given node, providing a measure of the node's connectivity. Next, we developed a function to compute the size of the largest connected component, taking into account the weights of the edges. This function identifies the largest weakly connected component in the graph and sums the weighted degrees of all nodes within this component. We then created a function to iteratively remove edges and calculate the weighted LCC size after each removal. It removes each edge in sequence, recalculates the weighted LCC size, and records the relative amount of removed edges along with the current weighted LCC size. This allows us to assess the impact of removing each edge on the overall connectivity of the network. To visualize the results, we generated plots that display the weighted LCC size against the relative amount of removed edges for multiple approaches. Each approach is represented by a different line on the plot, facilitating a comparative analysis of the algorithms' performance.

We analyzed the effectiveness of various algorithms in distinguishing between biological networks by comparing the density distributions of curvature values generated by each algorithm. To ensure a

fair comparison, we applied min-max normalization to the curvature values, scaling them to a common range of [0, 1]. This normalization step was crucial in eliminating the influence of different value ranges, allowing us to focus on the shape and distributional characteristics of the normalized data. We employed the Cramer-von Mises distance [79], the Wasserstein distance [80], and the Kolmogorov-Smirnov test [81] to evaluate the differences between the distributions produced by each algorithm. The Cramer-von Mises distance measures the cumulative squared differences between the empirical cumulative distribution function (ECDF) of two distributions. This metric provides a comprehensive assessment of the overall distributional differences, making it sensitive to variations across the entire range of data. The Wasserstein distance calculates the minimum amount of work required to transform one probability distribution into another by moving distribution mass, where the "work" is defined as the amount of distribution mass times the distance it is moved. The Kolmogorov-Smirnov test measures the maximum distance between the ECDFs of two distributions, highlighting the most significant point of divergence between the distributions.

Utilized Networks

Artificial Network

Two different types of random networks, generated with networkx [82], were used in this study to evaluate the performance of various network analysis algorithms. These networks were generated using specific scripts to create directed graphs with randomly assigned weights to both edges and nodes. Below are detailed descriptions of each network model used:

Small Random Network. A small random network with 10 nodes was created, and each possible directed edge between nodes was included with a probability of 0.3. To this network, random weights were assigned to the edges and nodes. The edge weights were uniformly distributed between 0.1 and 1.0, while the node weights were uniformly distributed between 0.1 and 1.0. This small random network served as a basic model to test the algorithms on a relatively simple and small-scale network.

Medium Random Network. Similar to the small random network, a larger random network was generated, consisting of 100 nodes, with each possible directed edge included with a probability of 0.3. Random weights were assigned to both edges and nodes. The edge weights were uniformly distributed between 0.1 and 1.0, and the node weights were uniformly distributed between 0.1 and 1.0. The larger size of this network provided a more complex structure to evaluate the algorithms on a medium-scale network.

These network models allowed for comprehensive testing and comparison of the network analysis algorithms under different conditions and complexities, enhancing the robustness and generalizability of the study's findings.

Biological Networks

E.coli under glucose and fucose growth conditions

We obtained transcriptomic and metabolomic data from *Escherichia coli* grown under two different conditions: once under glucose- and once under fucose-rich conditions. Transcriptomes and metabolomes were measured during exponential growth [72]. Genes were used as edge weights based on normalized transcriptomic values, by assigning normalized gene counts according to gene-reaction rules using Cobrapy [83]. Counts for isoenzymes were summed, while the lowest gene count was used for enzyme complexes. Every time after adding transcriptomic constraints to the iML1515 GSMM, we performed an FBA to ensure feasibility. In case of an infeasible solution, we avoided this reaction constraint. To include node weights for metabolites, we accounted for the extreme discrepancies in concentrations, such as the high concentrations of cofactors compared to other metabolites. This was addressed by performing a quantile transformation, assuming a normal distribution of metabolites

within a cell, using the `QuantileTransformer` function from `scikit-learn` [78]. The general procedure involved ranking the individual data points and then assigning values according to an underlying normal distribution. This transformation reduced the differences across metabolite concentrations from eight orders of magnitude to one order of magnitude. After the quantile transformation, the resulting values were used as node weights. Metabolites that were not measured were excluded from the analysis and automatically assigned a weight of 0.1. Additionally, intracellular and extracellular metabolites were transformed separately to ensure accuracy. This approach allowed for a more balanced and representative analysis of the metabolomic data under different growth conditions.

Microbial network

Levy, et al. investigated the difference of obesity in inflammatory bowel disease (IBD)-patients based on metabolic complementary and co-occurrence of 154 microbial species [1]. Pairwise metabolic complementarity values were obtained for each pair of microbes from [1], with values ranging from 0 to 1. Additionally, co-occurrence values for each species were obtained using Jaccard similarity coefficients under 4 different conditions: healthy–lean, healthy–obese, IBD–lean, and IBD–obese [1]. From these data, we constructed a network for each condition where each species was represented as a node, and the directed edges between nodes corresponded to the metabolic complementarity values. The node weights for each species were determined by summing the Jaccard similarity coefficients from the co-occurrence data for each species. This network structure allowed for the examination of microbial interactions and the role of metabolic complementarity in the microbiome.

Results and Discussion

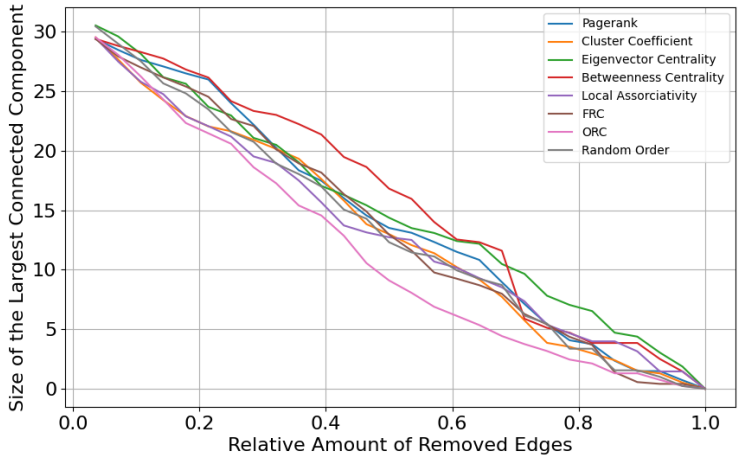
Random Networks

In an initial evaluation, we compared the edge ranking algorithms using a small network (Tab 1) and a medium-sized random network (Tab 2). These networks consisted of 10 and 100 nodes respectively, with 20 and 2934 edges respectively. When analyzing the weighted LCC of the network after sequentially removing edges according to their rankings, ORC demonstrated superior performance in both small and medium-sized networks when employing the original algorithms (Fig 2a, and Fig 4a). After normalizing the edge number, the FRC's performance closely matched that of the ORC in the weighted LCC plots, suggesting that edge number normalization might be a crucial enhancement for the FRC (Fig 3a, and Fig 5a). While most other algorithms (edge cluster coefficient, edge eigenvector centrality, and edge local assortativity) were indistinguishable from random edge removal, edge pagerank performed nearly as well as ORC without edge number normalization and surpassed FRC (Fig 2a, Fig 4a). However, upon applying edge number normalization, edge pagerank grouped with the other algorithms, whereas FRC performed comparably to ORC (Fig 3a, Fig 5a). When comparing the edge rankings identified by each method using PCA, clusters emerged with edge cluster coefficient and edge eigenvector centrality, while edge pagerank, FRC, and ORC formed another cluster without augmentation. Upon performing edge number normalization, edge pagerank clustered with edge eigenvector centrality and edge cluster coefficient, whereas ORC and FRC remained together (Fig 4b, Fig 5b). This clustering pattern was not observed in the smaller network, likely due to the insufficient number of edges, irrespective of edge number normalization (Fig 2b, Fig 3b). Across both approaches and networks, edge betweenness consistently performed the worst, even underperforming random edge removal (Fig 2a, Fig 3a, Fig 4a, Fig 5a). In terms of runtime, the fastest algorithms for the small network were edge local assortativity and edge betweenness centrality, while edge eigenvector centrality and edge pagerank outperformed others in the medium-sized network (Tab 2, Tab 3), regardless of edge number normalization. For both algorithms and networks, FRC and ORC exhibited the longest runtimes, nearly three orders of magnitude slower than the top-performing algorithms, which may present a potential disadvantage for larger networks.

Table 1. Network metrices of a small random network. To compare the performance and implementation of the edge ranking algorithms, a small random network was generated.

Metric	
Number of nodes	10
Number of edges	20
Average degree	2.80
Density	0.3111
Highest degree	7
Lowest degree	4

(a)



(b)

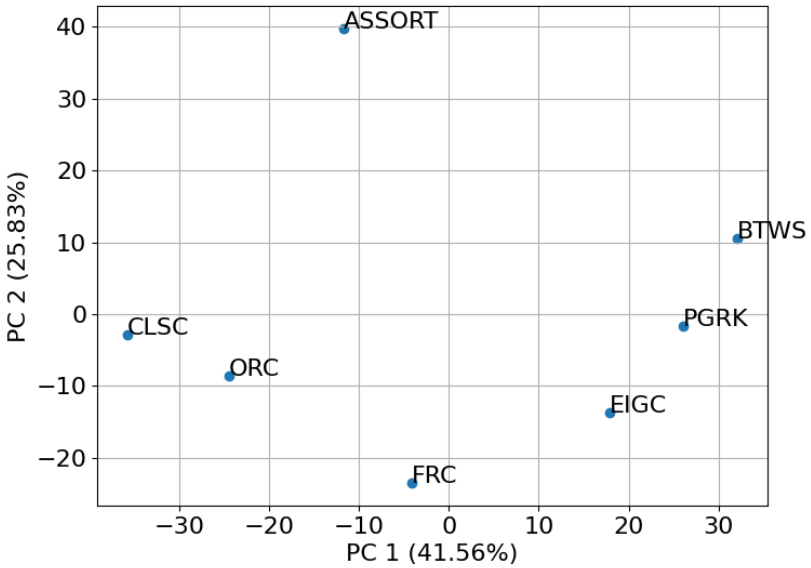
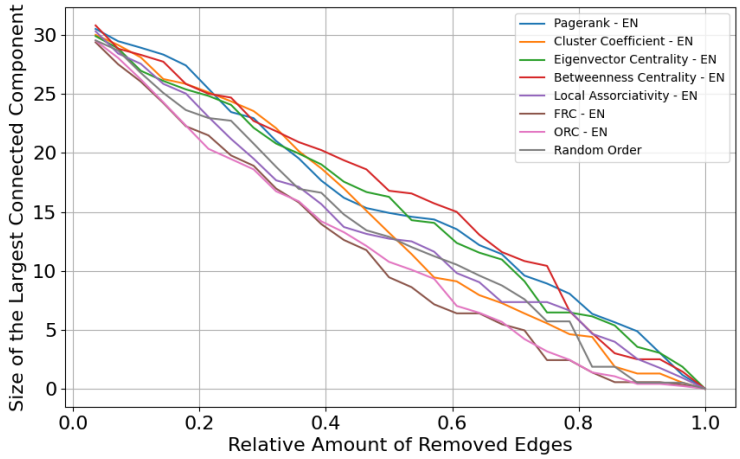


Fig 2. Comparison of seven edge ranking algorithms using a small random network. (a) Comparison of seven edge ranking algorithms based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS: edge betwenness, ASSORT: edge local assortiativity, PGRK: edge pagerank, EIGC: edge eigenvector centrality, CLSC: edge cluster coefficient, ORC: Ollivier-Ricci Curvature, FRC: Forman-Ricci Curvature.

(a)



(b)

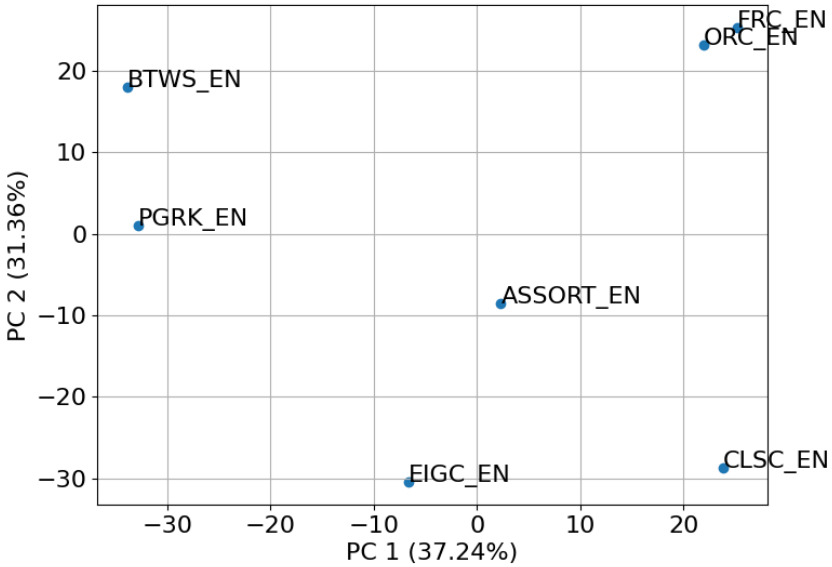


Fig 3. Comparison of seven edge ranking algorithms considering the normalization for edge number using a small random network. (a) Comparison of seven edge ranking algorithms with normalization for edge number based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking with normalization for edge number as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS_EN: edge betwenness, ASSORT_EN: edge local assortiativity, PGRK_EN: edge pagerank, EIGC_EN: edge eigenvector centrality, CLSC_EN: edge cluster coefficient, ORC_EN: Ollivier-Ricci Curvature, FRC_EN: Forman-Ricci Curvature, EN: normalized for edge number.

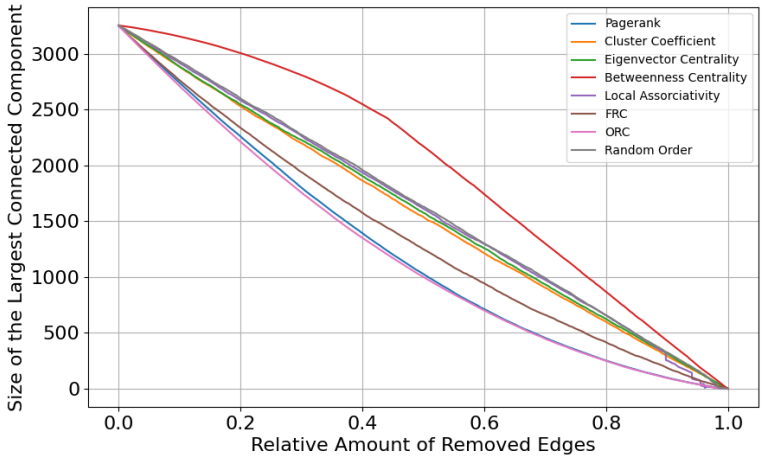
Table 2. Runtimes of algorithms applied to a small random network. The runtime was measured for each algorithm, when using the original implementation (Original), and the augmented algorithm when normalizing for edge numbers (EN). The fastest algorithm was edge local assortiativity, followed by edge betweenness, while the slowest algorithm were FRC and ORC.

Algorithm / Augmentation	Original	EN
Egde Betweenneess	0.0009	0.0008
Edge Pagerank	0.0024	0.0025
Edge Eigenvector Centrality	0.0013	0.0015
Edge Cluster Coefficient	0.0016	0.0015
Edge Local Assorciativity	0.0005	0.0005
Forman-Ricci-Curvature	0.0026	0.0025
Ollivier-Ricci-Curvature	0.0332	0.0328

Table 3. Network metrices of a medium-sized random network. To compare the performance and implementation of the edge ranking algorithms, a medium-sized random network was generated.

Metric	
Number of nodes	100
Number of edges	2934
Average degree	29.34
Density	0.2964
Highest degree	78
Lowest degree	37

(a)



(b)

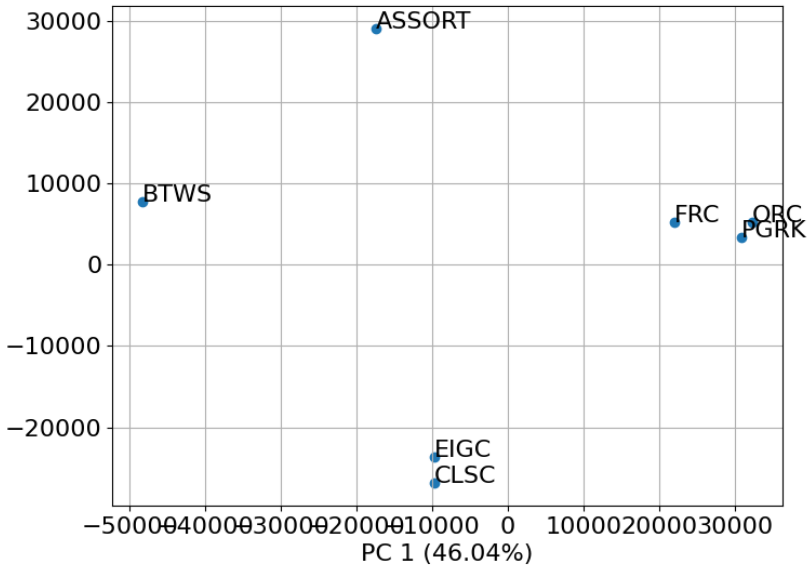
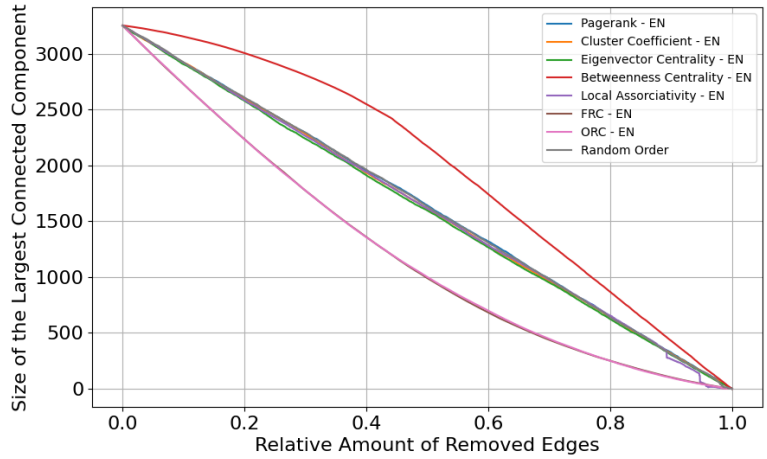


Fig 4. Comparison of seven edge ranking algorithms using a medium random network. (a) Comparison of seven edge ranking algorithms based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS: edge betwenness, ASSORT: edge local assortiativity, PGRK: edge pagerank, EIGC: edge eigenvector centrality, CLSC: edge cluster coefficient, ORC: Ollivier-Ricci Curvature, FRC: Forman-Ricci Curvature.

(a)



(b)

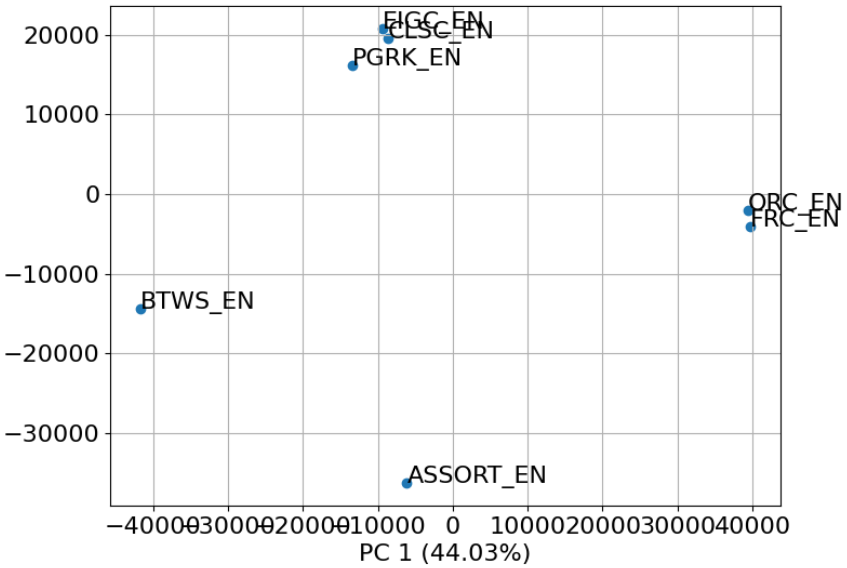


Fig 5. Comparison of seven edge ranking algorithms considering the normalization for edge number using a medium random network. (a) Comparison of seven edge ranking algorithms with normalization for edge number based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking with normalization for edge number as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS_EN: edge betwenness, ASSORT_EN: edge local assortiativity, PGRK_EN: edge pagerank, EIGC_EN: edge eigenvector centrality, CLSC_EN: edge cluster coefficient, ORC_EN: Ollivier-Ricci Curvature, FRC_EN: Forman-Ricci Curvature, EN: normalized for edge number.

Table 4. Runtimes of algorithms applied to a medium-sized random network. The runtime was measured for each algorithm, when using the original implementation (Original), and the augmented algorithm when normalizing for edge numbers (EN). The fastest algorithms were edge eigenvector centrality, and edge pagerank, while the slowest algorithms were FRC and ORC.

Algorithm / Augmentation	Original	EN
Egde Betweenneess	0.19	0.20
Edge Pagerank	0.02	0.01
Edge Eigenvector Centrality	0.01	0.01
Edge Cluster Coefficient	0.58	0.59
Edge Local Assorciativity	0.12	0.12
Forman-Ricci-Curvature	9.39	9.52
Ollivier-Ricci-Curvature	10.46	10.48

E.coli under glucose and fucose growth conditions

Next, we converted the GSMM iML1515 [84] into a metabolic graph, with metabolites as nodes and reactions as edges. We obtained two metabolic graphs from transcriptomic and metabolomic data during the exponential phase — one under glucose-rich media and one in fucose-rich media [72]. The reconstructed metabolic graph comprised 1877 nodes and 8599 directed edges. As is commonly known, the node degrees varied across the networks, with the highest node degree being 1097 and the lowest node degree being 1 (Tab 5). Similar to the medium-sized random network, ORC performed best when comparing the weighted LCC after removing ranked edges sequentially, followed by edge pagerank, edge cluster coefficient, and FRC (Fig 6a). In the PCA plot (Fig 6b), all algorithms except edge betweenness centrality formed a cluster, while FRC tended to cluster more closely with edge eigenvector centrality and edge local assortiativity. When considering edge number normalization in all algorithms (Fig 7), FRC performed almost as well as ORC, while the other algorithms maintained their performance or showed worse results. Although the size of the weighted LCC after removing the top 40% ranked edges was slightly smaller for ORC following edge number normalization, the edge number normalized FRC (FRC_EN) resulted in improved performance, with a weighted LCC of 1000 after removing 40% of the top-ranked edges (Fig 7a). This performance surpassed that of edge pagerank and edge cluster coefficient without edge number normalization (Fig 6a), where a weighted LCC of around 2000 was observed after removing the top 40% edges. The edge rankings also revealed a different clustering pattern in the PCA when accounting for edge number normalization, with FRC and ORC clustering closely together, while all other algorithms were more loosely distributed (Fig 7b). To determine if this result was network-specific, we conducted the same experiments on the network obtained from the metabolic model under glucose-rich conditions. As seen in Fig 8 and Fig 9, the trends were consistent with those observed in the fucose-rich network.

To further evaluate the performance of each algorithm in both their original implementation and after edge number normalization, we analyzed the glucose- and fucose-derived networks with respect to the distribution of edge-specific values for each algorithm (Fig 10h). While all algorithms displayed a peak, suggesting that the majority of edges in the network possess similar values, deviations were noted in certain distributions. For example, the edge eigenvector centrality distribution for the fucose-derived network exhibited lower frequencies at values around 0.2, whereas the glucose-derived network demonstrated a peak within this range (Fig 10c). Additionally, the FRC distribution for the fucose network shifted towards more negative values compared to the glucose network, indicating reduced stability in the fucose network due to these negative values (Fig 10f). Given that each algorithm generates edge-specific values on different scales, we normalized these scales as outlined in the methods section to assess whether this affected the differentiation between the networks. Although minor changes were observed for edge betweenness (Fig 11a) and ORC (Fig 11g), the normalization of edge-specific values did not substantially alter the separation between the two networks (Fig 11h). Subsequently, we examined edge-specific distributions from both networks following edge number normalization of values (Fig 13h, and Fig 14h). The edge number normalized distributions maintained similar shapes to those observed without normalization, except for the FRC (Fig 13f). In this case, the fucose network included some edges with higher values compared to the glucose network. Utilizing normalized distributions of the edge number normalized values revealed a distinct separation between the two networks for the FRC (Fig 14f), while all other methods resulted in similar distributions (Fig 14h).

Next, we evaluated the performance of the algorithms in distinguishing between the fucose and glucose networks using three statistical measures to compare distributions: Cramer-von Mises distance, Wasserstein distance, and the Kolmogorov-Smirnov test, as shown in Fig 12 and Fig 15. Focusing on the original algorithms, we found that edge eigenvector centrality and edge cluster coefficient provided the best separation according to the Cramer-von Mises distance (Fig 12a) and the Kolmogorov-Smirnov test (Fig 12e). However, when assessing the Wasserstein distance (Fig 12c), ORC and FRC demonstrated the highest performance. In contrast, when using normalized distributions, FRC outperformed the other algorithms based on the Cramer-von Mises distance (Fig 12b) and the Kolmogorov-Smirnov test (Fig 12f), while edge eigenvector centrality was the most

effective for distinguishing the networks using the Wasserstein distance (Fig 12d). Considering the edge number normalization of edge-specific values, edge eigenvector centrality again emerged as the best algorithm for separation based on the Cramer-von Mises distance (Fig 15a) and the Kolmogorov-Smirnov test (Fig 15c), whereas FRC was superior when applying the Wasserstein distance (Fig 15c). Moreover, when normalizing the distributions of edge number normalized algorithms, FRC consistently outperformed all other algorithms, regardless of the statistical measure applied (Fig 15b,d,f).

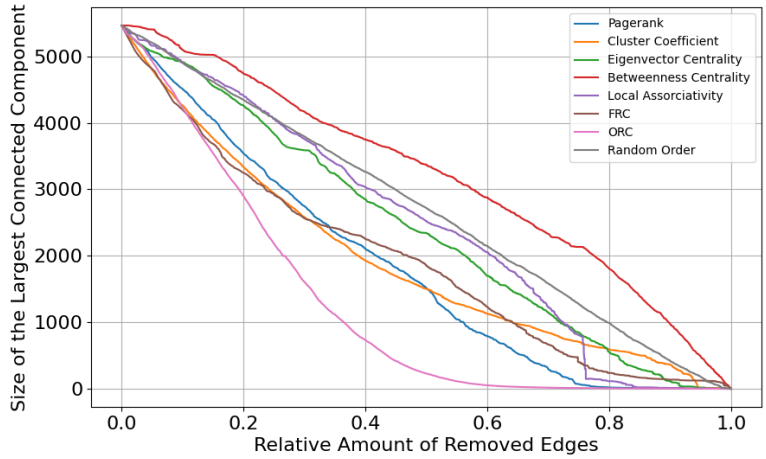
To further elucidate the capabilities of the algorithms in identifying specific targets, we extracted the top 5 ranked edges from each algorithm and compared them between the two networks, as well as with edge number normalization, as shown in Tab 6. In both fucose and glucose networks, the top 5 edges identified by edge betweenness, edge eigenvector centrality, and edge local assortativity were cofactor-associated reactions and were similar across both networks, making it challenging to draw meaningful conclusions. The top 5 ranked edges from edge pagerank and ORC were identical for both the fucose and glucose networks. For FRC, the top 5 were also the same except for one edge, which involved the conversion of mannose to mannose-6-phosphate. This edge was significant in the fucose network but not in the glucose network. According to KEGG [85], there is no direct link between fucose metabolism and mannose metabolism. The edge cluster coefficient allowed a more detailed comparison of the top 5 ranked edges: while edges involved in curcumin metabolism were ranked higher in the fucose network, edges involved in peptidoglycan synthesis and purine metabolism were ranked higher in the glucose network. It is reasonable that under glucose-rich conditions, nucleotide and peptidoglycan synthesis become more crucial for growth [86, 87], and therefore these edges are potentially essential. However, there is no information in the literature that suggests curcumin metabolism is enriched under fucose-rich conditions. When considering edge number normalization, cofactor-associated edges were again top-ranked in both networks for edge betweenness, edge pagerank, edge eigenvector centrality, and edge local assortativity. The edges identified by ORC after edge number normalization were similar to those identified by the original ORC. Using edge cluster coefficient, curcumin metabolism edges remained top-ranked in the fucose network, while edges involved in converting various lipids into acyl phosphatidylglycerol (n16:1) were top-ranked in the glucose network. With FRC and edge number normalization, significant differences were observed between the fucose and glucose networks among the top-ranked edges: in the fucose network, the import of various diacylglycerols from the periplasm into the cytosol was important, whereas in the glucose network, edges associated with Kdo2-lipidA and peptidoglycan synthesis were top-ranked. This can be explained by the fact that ATP is limiting under fucose conditions [72], necessitating the import of lipids to meet energy demands. Conversely, in glucose-rich conditions, the synthesis of complex peptidoglycan or Kdo2-lipidA might be a rate-limiting step for growth, making these edges essential in the network.

Finally, we compared the runtimes for each algorithm once more and observed that this time FRC had the longest runtime, taking nearly twice as long as ORC and being 4 orders of magnitude slower than the fastest algorithm (Tab 7).

Table 5. Network metrices of a metabolic graph, derived from the GSMM iML1515. The metabolic graphs for fucose- and glucose specific conditions had identical metrices, since they were derived from the same GSMM, but with different node and edge weights, as explained in the methods.

Metric	
Number of nodes	1877
Number of edges	8599
Average degree	4.58
Density	0.0024
Highest degree	1097
Lowest degree	1

(a)



(b)

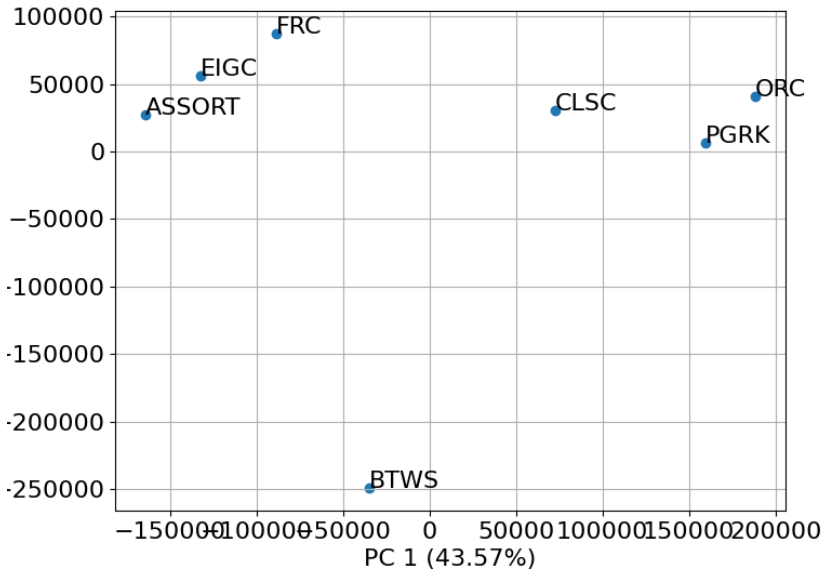
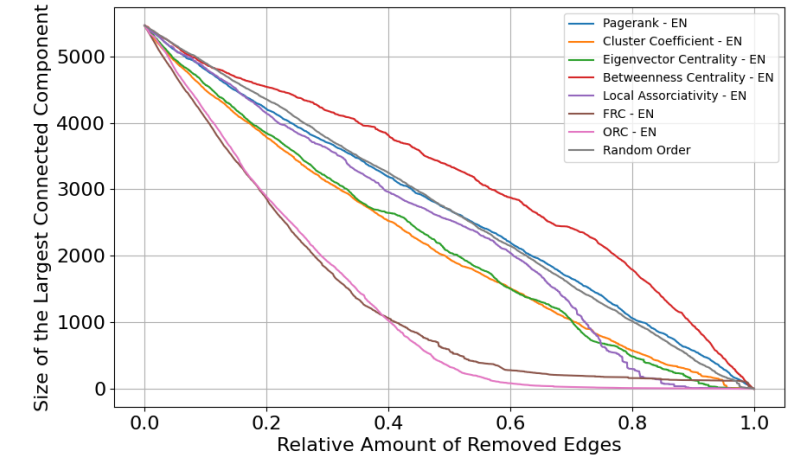


Fig 6. Comparison of seven edge ranking algorithms using the metabolic graph derived from the GSMM iML1515 under fucose-rich condition. (a) Comparison of seven edge ranking algorithms based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS: edge betwenness, ASSORT: edge local assortiativity, PGRK: edge pagerank, EIGC: edge eigenvector centrality, CLSC: edge cluster coefficient, ORC: Ollivier-Ricci Curvature, FRC: Forman-Ricci Curvature.

(a)



(b)

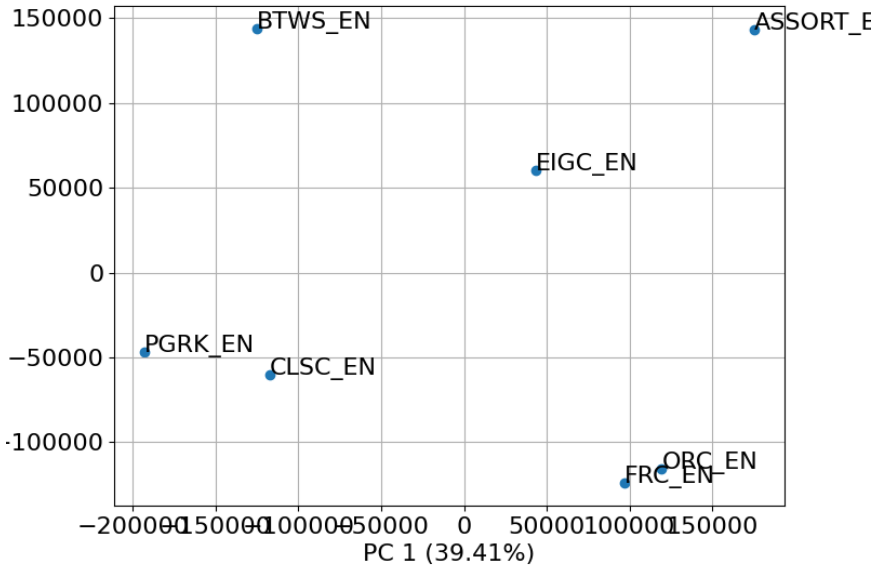
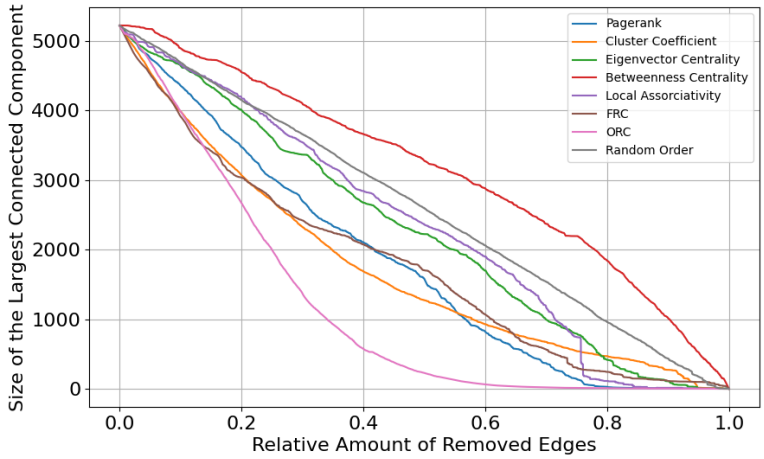


Fig 7. Comparison of seven edge ranking algorithms considering the normalization for edge number using the metabolic graph derived from the GSMM iML1515 under fucose-rich condition. (a) Comparison of seven edge ranking algorithms with normalization for edge number based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking with normalization for edge number as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS_EN: edge betwenness, ASSORT_EN: edge local assortiativity, PGRK_EN: edge pagerank, EIGC_EN: edge eigenvector centrality, CLSC_EN: edge cluster coefficient, ORC_EN: Ollivier-Ricci Curvature, FRC_EN: Forman-Ricci Curvature, EN: normalized for edge number.

(a)



(b)

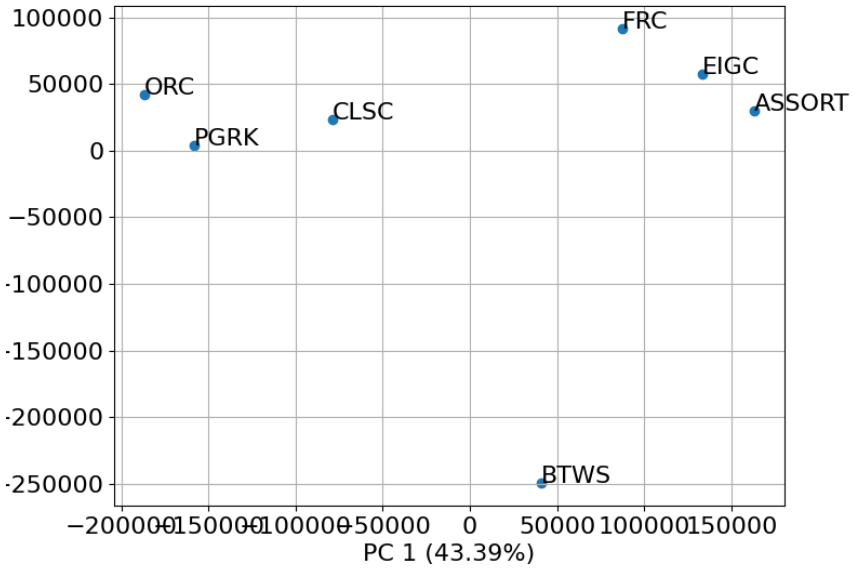
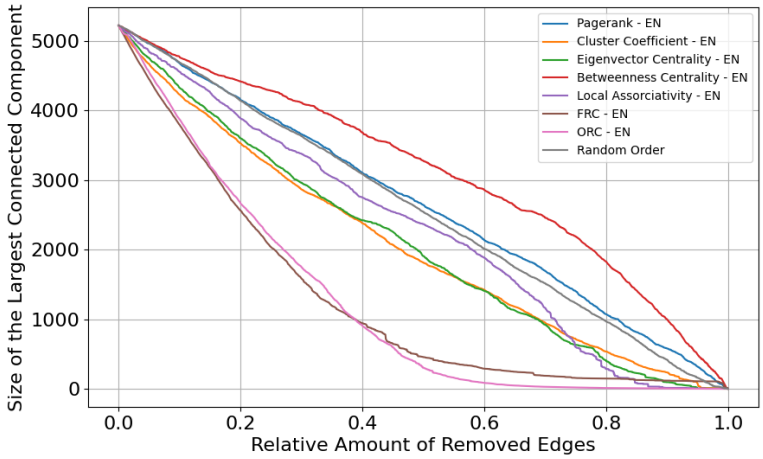


Fig 8. Comparison of seven edge ranking algorithms using the metabolic graph derived from the GSMM iML1515 under glucose-rich condition. (a) Comparison of seven edge ranking algorithms based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS: edge betwenness, ASSORT: edge local assortiativity, PGRK: edge pagerank, EIGC: edge eigenvector centrality, CLSC: edge cluster coefficient, ORC: Ollivier-Ricci Curvature, FRC: Forman-Ricci Curvature.

(a)



(b)

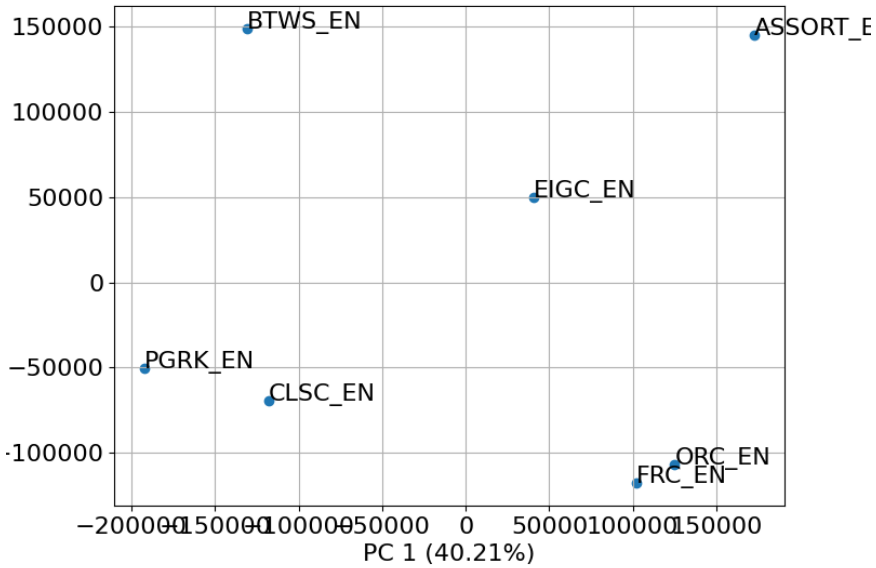


Fig 9. Comparison of seven edge ranking algorithms considering the normalization for edge number using the metabolic graph derived from the GSMM iML1515 under glucose-rich condition. (a) Comparison of seven edge ranking algorithms with normalization for edge number based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking with normalization for edge number as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS_EN: edge betwenness, ASSORT_EN: edge local assortiativity, PGRK_EN: edge pagerank, EIGC_EN: edge eigenvector centrality, CLSC_EN: edge cluster coefficient, ORC_EN: Ollivier-Ricci Curvature, FRC_EN: Forman-Ricci Curvature, EN: normalized for edge number.

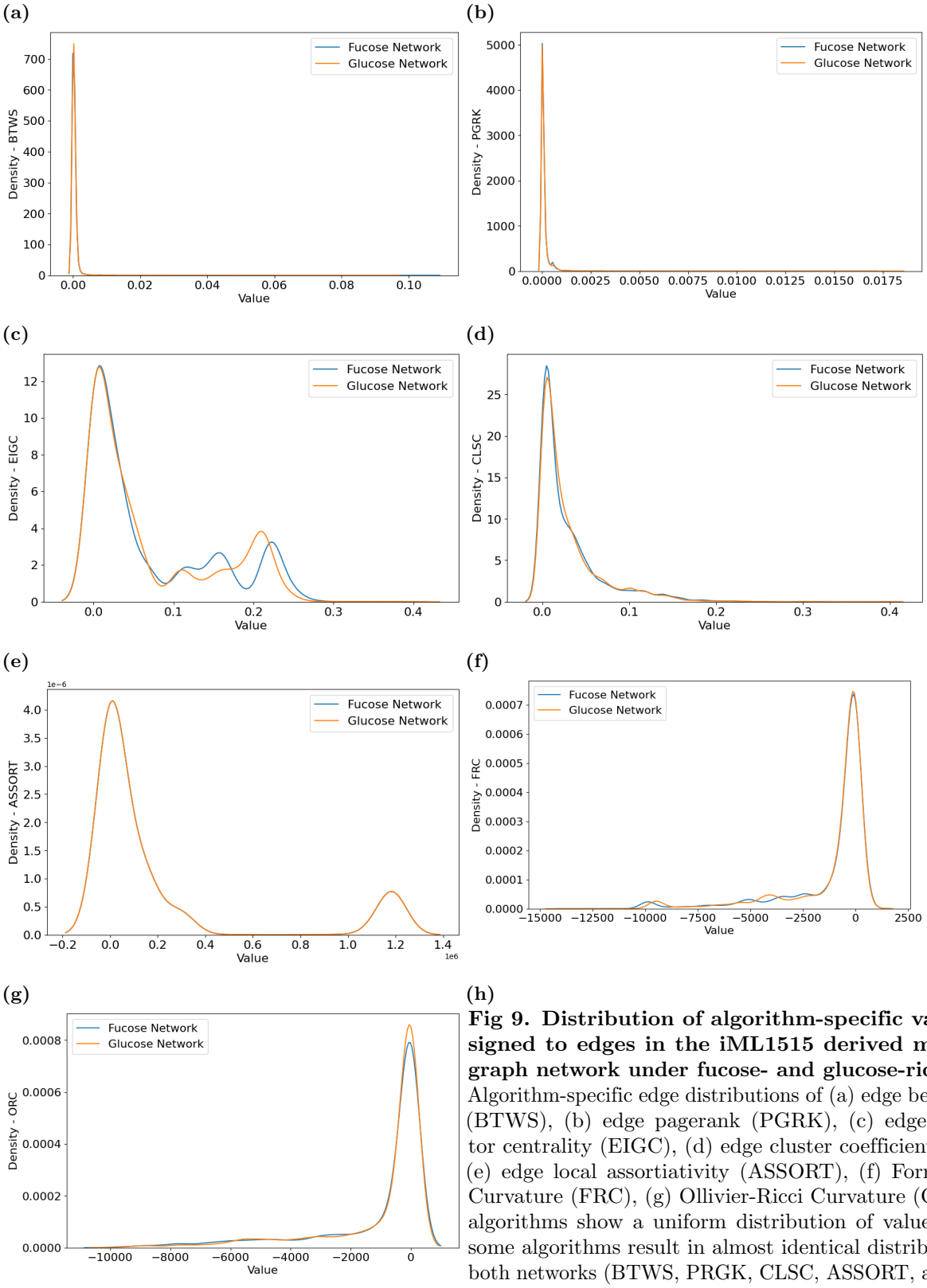


Fig 9. Distribution of algorithm-specific values, assigned to edges in the iML1515 derived metabolic graph network under fucose- and glucose-rich media. Algorithm-specific edge distributions of (a) edge betweenness (BTWS), (b) edge pagerank (PGRK), (c) edge eigenvector centrality (EIGC), (d) edge cluster coefficient (CLSC), (e) edge local assortiativity (ASSORT), (f) Forman-Ricci Curvature (FRC), (g) Ollivier-Ricci Curvature (ORC). All algorithms show a uniform distribution of values. While some algorithms result in almost identical distributions for both networks (BTWS, PRGK, CLSC, ASSORT, and ORC), two algorithms (EIGC, and FRC) allow for a differentiation between the two networks, when comparing the more important edges aside from the main peak.

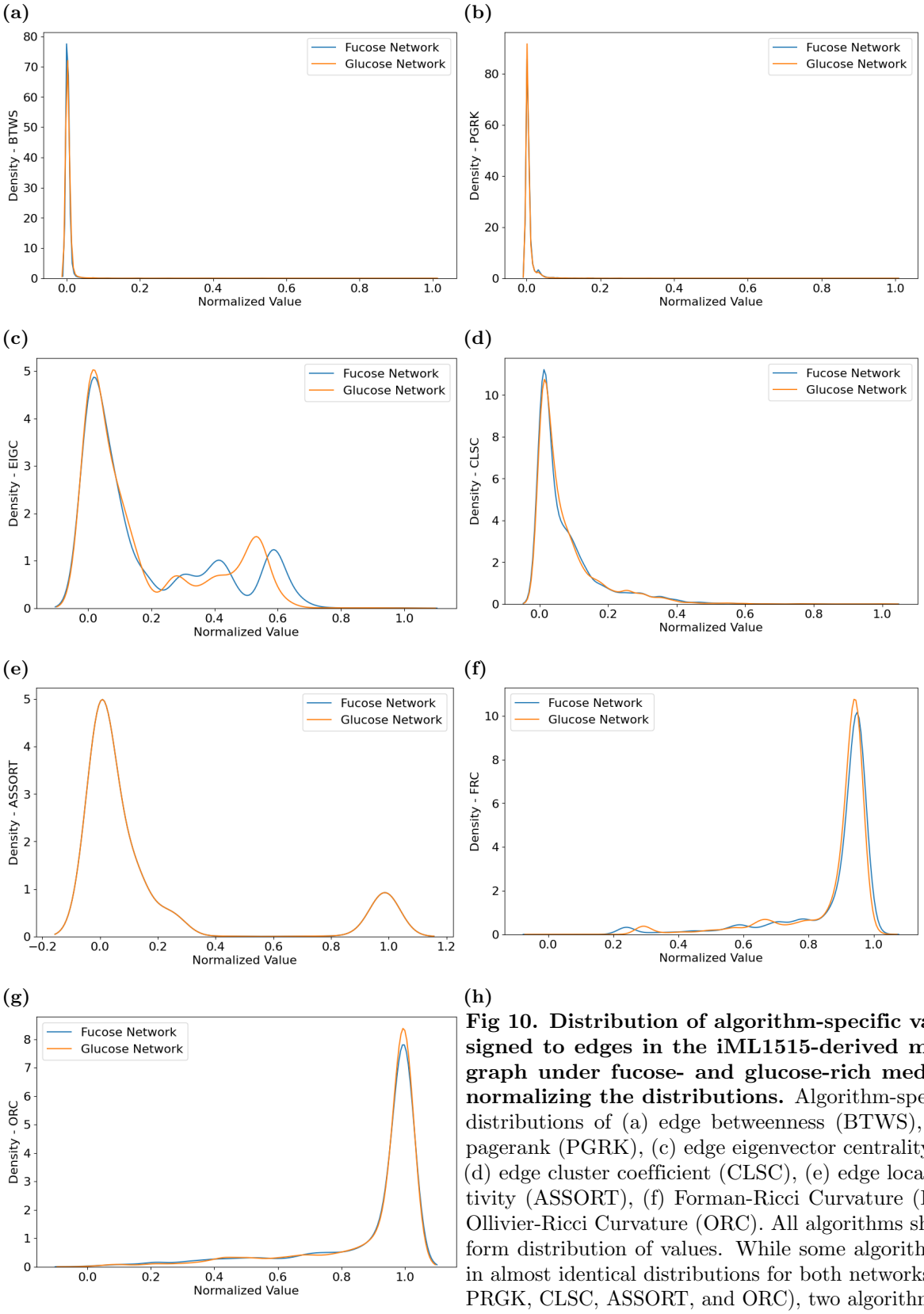


Fig 10. Distribution of algorithm-specific values, assigned to edges in the iML1515-derived metabolic graph under fucose- and glucose-rich media when normalizing the distributions. Algorithm-specific edge distributions of (a) edge betweenness (BTWS), (b) edge pagerank (PGRK), (c) edge eigenvector centrality (EIGC), (d) edge cluster coefficient (CLSC), (e) edge local assortativity (ASSORT), (f) Forman-Ricci Curvature (FRC), (g) Ollivier-Ricci Curvature (ORC). All algorithms show a uniform distribution of values. While some algorithms result in almost identical distributions for both networks (BTWS, PRGK, CLSC, ASSORT, and ORC), two algorithms (EIGC, and FRC) allow for a differentiation between the two networks, when comparing the more important edges aside from the main peak.

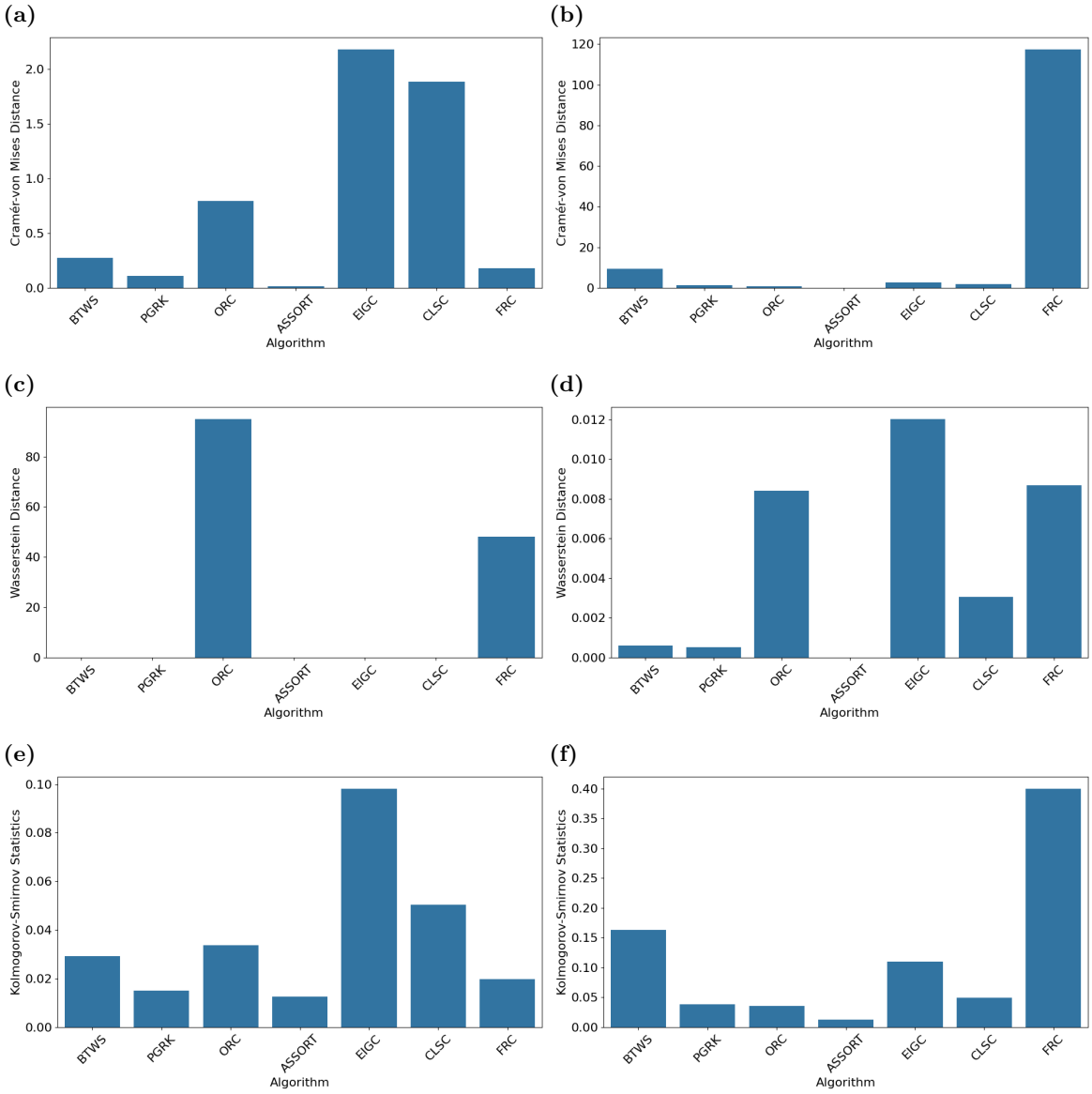


Fig 11. Evaluation of edge-ranking algorithms algorithms and their capability to differentiate between networks. Cramer-von Mises distance (a, b), Wasserstein distance (c, d), and Kolmogorov-Smirnow test (e, f) to identify which algorithm captures the difference of metabolic graphs, derived from fucose- and glucose-rich conditions better, when considering original distributions (a, c, e), or normalized distributions (b, d, f).

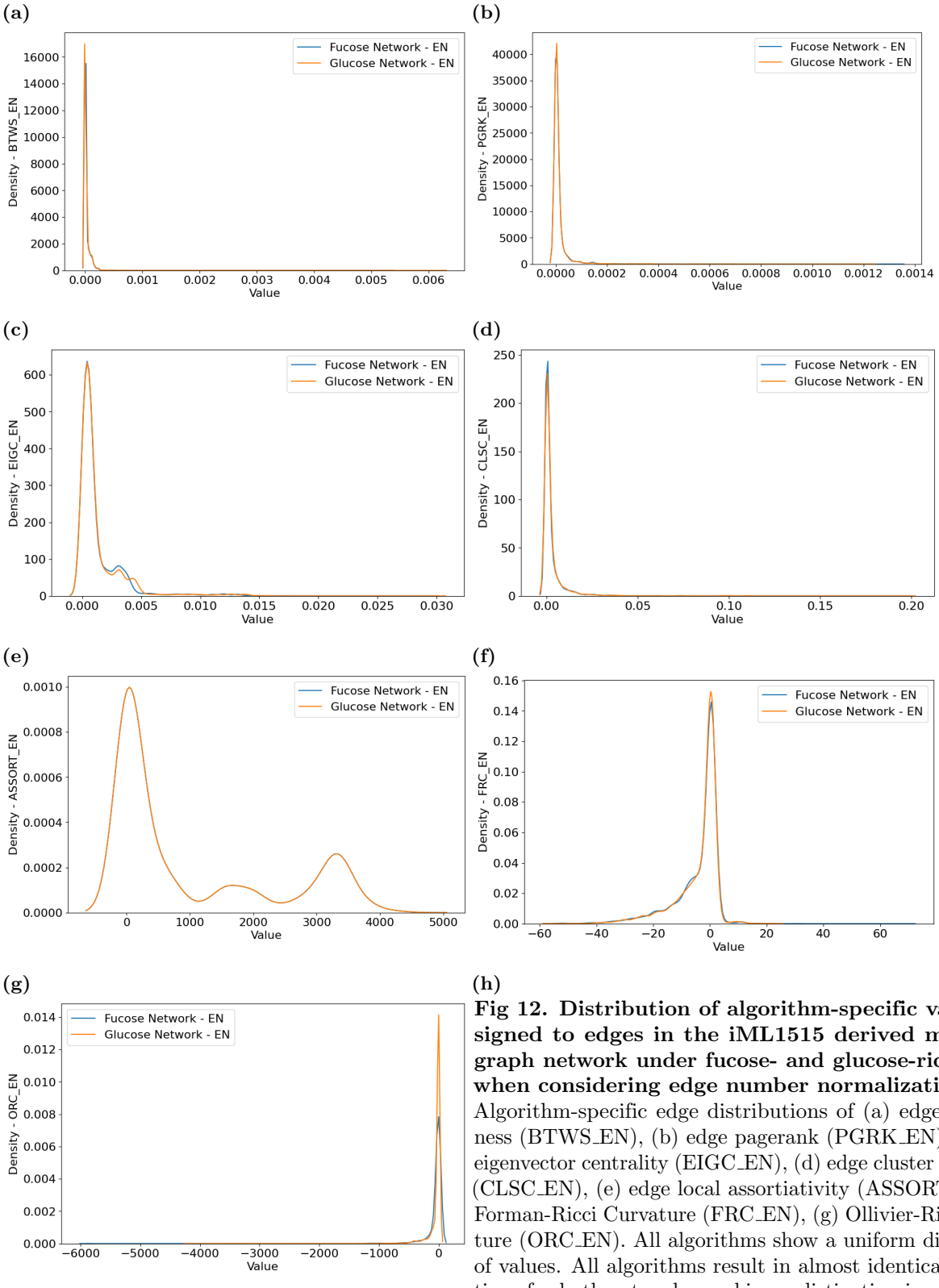


Fig 12. Distribution of algorithm-specific values, assigned to edges in the iML1515 derived metabolic graph network under fucose- and glucose-rich media, when considering edge number normalization (EN). Algorithm-specific edge distributions of (a) edge betweenness (BTWS_EN), (b) edge pagerank (PGRK_EN), (c) edge eigenvector centrality (EIGC_EN), (d) edge cluster coefficient (CLSC_EN), (e) edge local assortiativity (ASSORT_EN), (f) Forman-Ricci Curvature (FRC_EN), (g) Ollivier-Ricci Curvature (ORC_EN). All algorithms show a uniform distribution of values. All algorithms result in almost identical distributions for both networks, making a distinction impossible.

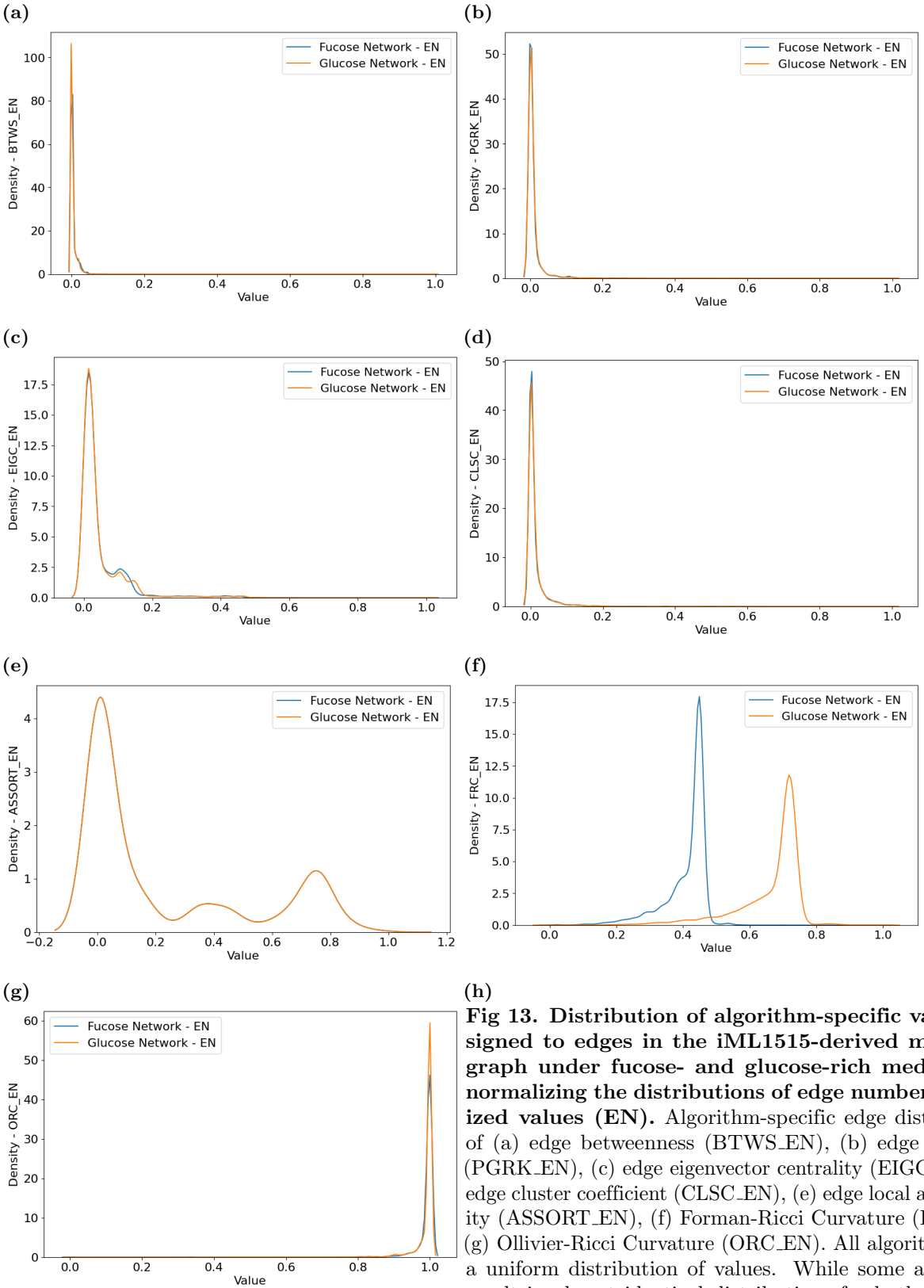


Fig 13. Distribution of algorithm-specific values, assigned to edges in the iML1515-derived metabolic graph under fucose- and glucose-rich media when normalizing the distributions of edge number normalized values (EN). Algorithm-specific edge distributions of (a) edge betweenness (BTWS_EN), (b) edge pagerank (PGRK_EN), (c) edge eigenvector centrality (EIGC_EN), (d) edge cluster coefficient (CLSC_EN), (e) edge local assortativity (ASSORT_EN), (f) Forman-Ricci Curvature (FRC_EN), (g) Ollivier-Ricci Curvature (ORC_EN). All algorithms show a uniform distribution of values. While some algorithms result in almost identical distributions for both networks, only FRC_EN allows for a differentiation between the two networks.

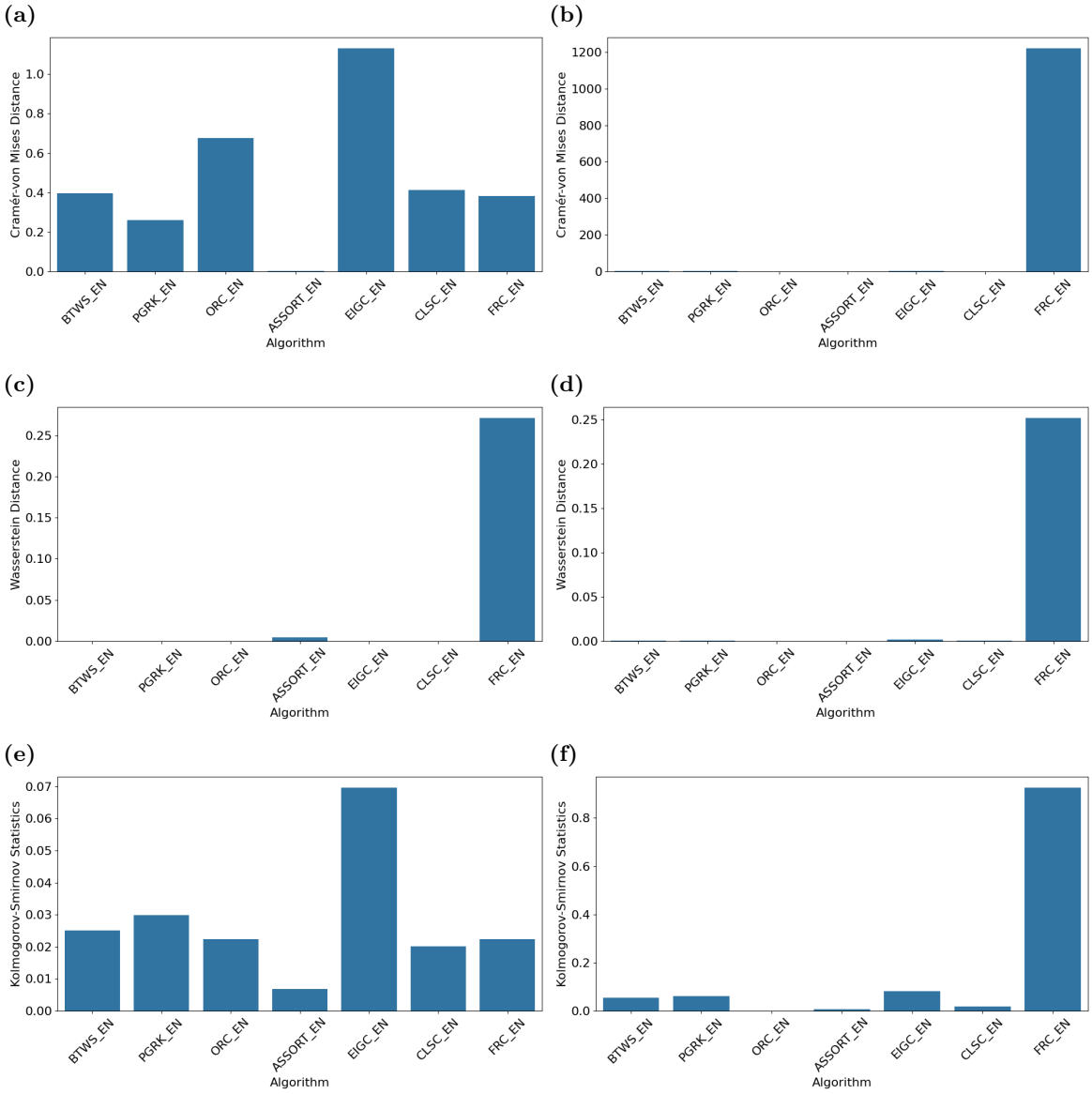


Fig 14. Evaluation of edge-ranking algorithms algorithms and their capability to differentiate between metabolic graphs when considering edge number normalization. Cramer-von Mises distance (a, b), Wasserstein distance (c, d), and Kolmogorov-Smirnow test (e, f) to identify which algorithm captures the difference of metabolic graphs, derived from fucose- and glucose-rich conditions better, when considering original distributions (a, c, e), or normalized distributions (b, d, f) and edge number normalization.

Table 6. Top 5 ranked edges obtained from each algorithm and network. Top-ranked edges from each algorithm for both networks were extracted considering the original implementation and edge number normalization (EN). When relying on the original implementations, most algorithms result in edges, that belong to cofactors, or are the same for both networks, making it difficult to identify differences. When considering EN-algorithms, it was possible to identify edges assigned to β oxidation in fucose networks, while lipidA-synthesis was found to be important in glucose networks, when using FRC_EN.

Algorithm	Fucose-graph	Glucose-graph	Fucose-graph (DN)	Glucose-graph (DN)
Edge Betweenness	*ade_c*, *h_p*	*h_c*, 10*fthf_c*	*glu1sa_c*, 5*aop_c*	*no2_p*, *no2_c*
	ppi_c, *atp_c*	10*fthf_c*, *h2o_c*	23*dhba_c*, *amp_c*	*dxyl5p_c*, *ppi_c*
	fe2_p, *h2o_p*	*h_c*, *no2_p*	23*ddhb_c*, 23*dhb_c*	*dtdp4aaddg_c*, *unagamuf_c*
	amp_c, *ade_c*	*no2_p*, *no2_c*	*dtdp4aaddg_c*, *unagamuf_c*	*ohpb_c*, 4*per_c*
	pi_c, *ppi_c*	*no2_c*, *h_p*	*fe2_c*, *fe2_p*	*glyc_c*, *dha_c*
Edge Pagerank	*ppi_c*, *nicrnt_c*	*ppi_c*, *nicrnt_c*	*thym_e*, *thym_p*	*thym_e*, *thym_p*
	dmpp_c, *ipdp_c*	*dmpp_c*, *ipdp_c*	3*hpp_e*, 3*hpp_p*	3*hpp_e*, 3*hpp_p*
	ipdp_c, *dmpp_c*	*ipdp_c*, *dmpp_c*	5*mtr_e*, 5*mtr_p*	5*mtr_e*, 5*mtr_p*
	nicrnt_c, *dnad_c*	*nicrnt_c*, *ppi_c*	*thym_p*, *thym_e*	*thym_p*, *thym_e*
	nicrnt_c, *ppi_c*	*nicrnt_c*, *dnad_c*	*tma_e*, *tma_p*	3*hpp_p*, 3*hpp_e*
Edge Eigenvector Centr.	*h_c*, *adp_c*	*h_c*, *adp_c*	*gdpofuc_c*, *gdpfuc_c*	*gdpofuc_c*, *gdpfuc_c*
	adp_c, *h_c*	*adp_c*, *h_c*	*thmpp_c*, *athtp_c*	*thmpp_c*, *athtp_c*
	h_c, *pi_c*	*pi_c*, *h_c*	*ditp_c*, *didp_c*	*ditp_c*, *didp_c*
	pi_c, *h_c*	*h_c*, *pi_c*	*xtp_c*, *xdp_c*	*xtp_c*, *xdp_c*
	ppi_c, *h_c*	*h_c*, *ppi_c*	*met__D_p*, *met__D_c*	*ditp_c*, *ppi_c*
Edge Cluster Coeff.	*thmpp_c*, *athtp_c*	*thmpp_c*, *athtp_c*	*thmpp_c*, *athtp_c*	*thmpp_c*, *athtp_c*
	cur_c, *dhcur_c*	*glu__D_c*, *uamag_c*	*met__D_p*, *met__D_c*	*cur_c*, *dhcur_c*
	dhcur_c, *thcur_c*	*ditp_c*, *didp_c*	*cur_c*, *dhcur_c*	2*agpg*161*_c*, *apg*161*_c*
	dgslnt_c, *gslnt_c*	*xtp_c*, *xdp_c*	*dhcur_c*, *thcur_c*	2*agpg*181*_c*, *apg*181*_c*
	ditp_c, *didp_c*	*dgslnt_c*, *gslnt_c*	*ditp_c*, *didp_c*	2*agpg*140*_c*, *apg*140*_c*
Edge Local Assortativity	*adocbl_e*, *h_c*	*fe*3*dhbzs_e*, *h_c*	*sucbz_c*, *ppi_c*	*sucbz_c*, *ppi_c*
	*cbl*1*_e*, *h_c*	*adocbl_e*, *h_c*	4*hbz_c*, *ppi_c*	4*hbz_c*, *ppi_c*
	cbi_e, *h_c*	*cbi_e*, *h_c*	*cobalt*2*_c*, *ppi_c*	*cobalt*2*_c*, *ppi_c*
	*fe*3*dhbzs_e*, *h_c*	*cbl*1*_e*, *h_c*	*octa_p*, *ppi_c*	*octa_p*, *ppi_c*
	ag_c, *h_c*	*gp*4*g_c*, *h_c*	*dca_p*, *ppi_c*	*dca_p*, *ppi_c*
Forman-Ricci Curvature	*dca_c*, *amp_c*	*wco_c*, *amp_c*	2*aobut_c*, *accoa_c*	*ggagicolipa_c*, *gggagicolipa_c*
	man_c, *man*6*p_c*	*tungs_c*, *amp_c*	12*dgr*161*_p*, 12*dgr*161*_c*	*phphhlipa_c*, *kphphhlipa_c*
	wco_c, *amp_c*	*mobd_c*, *amp_c*	12*dgr*181*_p*, 12*dgr*181*_c*	*pep_c*, *mnl*1*p_c*
	moco_c, *amp_c*	*mptamp_c*, *amp_c*	12*dgr*120*_p*, 12*dgr*120*_c*	*anhm*4*p_c*, *LalaDgluMdapDala_c*
	mobd_c, *amp_c*	*moco_c*, *amp_c*	12*dgr*140*_p*, 12*dgr*140*_c*	23*dhbzs_c*, 23*dhb_c*
Ollivier-Ricci Curvature	*nh*4*_c*, *nadp_c*	*nh*4*_c*, *nadp_c*	*pheme_c*, *hemeO_c*	*dnad_c*, *nicrnt_c*
	gar_c, *thf_c*	*glu__L_c*, *nh*4*_c*	*dnad_c*, *nicrnt_c*	*slnt_c*, *dgslnt_c*
	glu__L_c, *nh*4*_c*	*feenter_c*, *ribflv_c*	*mobd_c*, *cu*2*_c*	*anhm*4*p_c*, *anhm_c*
	dtdp_c, *dtmp_c*	*dnad_c*, *nicrnt_c*	*slnt_c*, *dgslnt_c*	*uamag_c*, *ugmd_c*
	mobd_c, *cu*2*_c*	*uri_p*, *uri_c*	*uamag_c*, *ugmd_c*	*feenter_c*, *ribflv_c*

Table 7. Runtimes of algorithms applied to metabolic graphs, derived from fucose- or glucose-rich conditions. The runtime was measured for each algorithm, when using the original implementation (Original), and the augmented algorithm when normalizing for edge number (EN). The fastest algorithm was edge pagerank, followed by edge eigenvector centrality, while the slowest algorithms were FRC and ORC.

Algorithm / Augmentation	Original	EN
Egde Betweenneess	17.58	17.75
Edge Pagerank	0.09	0.05
Edge Eigenvector Centrality	0.12	0.14
Edge Cluster Coefficient	0.68	0.67
Edge Local Assorciativity	0.60	0.61
Forman-Ricci-Curvature	159.07	158.23
Ollivier-Ricci-Curvature	82.94	82.79

Microbial network

The microbial dataset obtained from Levy et al. [1] consists of 154 microbes, represented as nodes, and 23,188 species-to-species interactions, represented as directed edges (Tab 8). Unlike the previously described networks, the microbial network exhibited a much higher density, as every species shared an edge with every other species. First, we compared the performance of the microbial networks across different algorithms, with and without edge number normalization (Fig 16 and Fig 17). In contrast to previous networks, FRC outperformed all other algorithms except ORC, regardless of edge number normalization, when focusing on weighted LCC curves (Fig 16a, and Fig 17a). However, after edge number normalization, FRC's performance improved further, closely matching the ORC curve. With edge number normalization, we observed a weighted LCC of 3000 after removing the top 40% of edges, while the original FRC showed a weighted LCC of approximately 3500 after removing the top 40% of edges. Additionally, the clustering of edge rankings was consistently similar between FRC and ORC, irrespective of edge number normalization (Fig 16b, and Fig 17b).

Next, we compared the distributions obtained from FRC across four different microbial networks under four distinct conditions. All other algorithms produced identical distributions, as they rely solely on edge weights, whereas FRC can simultaneously consider both edge and node weights. The edge weights remained constant across all networks because the metabolic complementation score did not vary between states. Only the node weights were adjusted based on co-occurrence scores obtained from [1]. As with previous networks, we analyzed the original distribution (Fig 18a), the distribution with edge number normalization (Fig 18b), and the normalized distributions for both approaches (Fig 18c, Fig 18d). Based on the original implementation of FRC, the networks from IBD patients who are obese showed the most pronounced shift towards negative values, indicating that this network is the least stable. Conversely, the networks from healthy individuals (both lean and obese) were quite similar, while the network from lean IBD patients showed a slight shift towards higher values, indicating greater stability within the network (Fig 18a). This trend persisted after normalization of the distribution, although the difference between the curves was reduced (Fig 18c). Interestingly, normalization for edge number resulted in better overlapping distributions, distinguishable more by their height than by their shift (Fig 18b). Finally, normalization of the distribution of edge number normalized FRC-values yielded four nearly identical curves, with minimal separation between states (Fig 18d).

Next, we analyzed which distributions exhibited the highest variance across states and approaches. We utilized Cramer-von Mises distance, Wasserstein distance, and Kolmogorov-Smirnov statistics, as shown in Fig 19 and Fig 20. Across all approaches and distributions, the greatest difference was observed between obese and lean patients suffering from IBD, indicating that the microbial network is most impaired in obese individuals with IBD. While the prevalence of obesity in IBD is increasing and obesity appears to negatively impact various aspects of IBD [88], the full effects are still being elucidated and require further study [89]. In this small-scale microbial network, we identified a significant change associated with obesity in IBD patients. The second largest difference in FRC distributions was observed between healthy and IBD patients with obesity (Fig 19). It has been shown that the microbiome in obese individuals is dysregulated [90], and this dysregulation is even more pronounced when obese patients suffer from IBD [91]. However, we did not observe a significant difference between obese and lean individuals, provided they are healthy. This lack of significant difference might be due to the limited number of species within the network. Normalization of the distributions yielded less insightful results, particularly for edge number normalized distributions, where the greatest difference was observed between healthy obese and IBD lean individuals (Fig 20b,d,f).

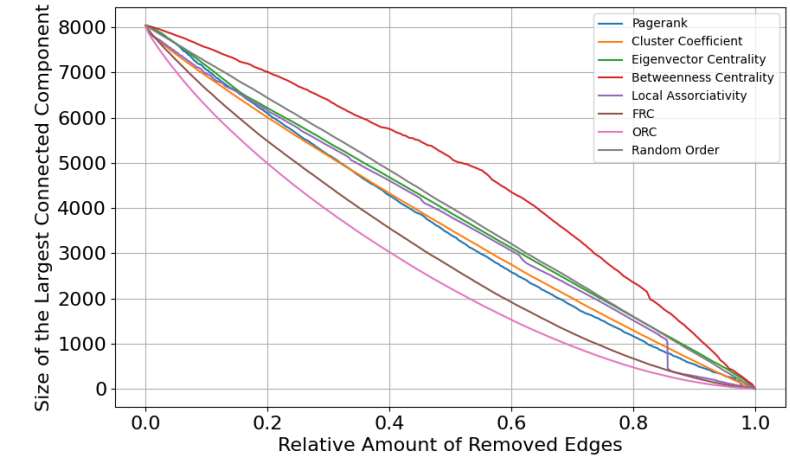
When focusing on the top 10 ranked edges from healthy and IBD patients, we identified various interactions as most crucial, as shown in Tab 9. In healthy, lean individuals, interactions primarily involved *Candidatus Sulcia muelleri* (Bacteriodes) with other bacteriodes among the top 10 ranked edges. However, in healthy, obese individuals, there were fewer such interactions. In this group, half of the interactions still involved *Candidatus Sulcia muelleri*, but the other half involved *Collinsella*

intestinalis [92,93]. The role of *Collinsella intestinalis* in IBD is unclear, with studies indicating it may be more abundant [92] or less abundant [94] in IBD patients. Among the top ten ranked edges, no *Collinsella intestinalis* strains were found in lean individuals with IBD. In contrast, obese individuals with IBD had half of the top-ranked edges containing *Collinsella intestinalis*, consistent with previous findings [92]. Additionally, *Gordonibacter pamelaeae* species were among the top-ranked edges in obese individuals with IBD. This bacterium has been implicated in the dysbiosis observed in IBD patients, raising questions about its role as an opportunistic pathogen or a problematic member of the normal flora in the context of inflammation [95,96]. Another notable difference in obese IBD individuals was that the second species involved in top-ranked edges were mostly bacteriodes, while the top four ranked edges included *Klebsiella pneumoniae*, *Escherichia fergusonii*, and *Salmonella enterica*, all belonging to Enterobacteriaceae. The reduction of bacteriodes species in IBD has been documented [97] and was also observed in networks analyzed using FRC. This separation among the top 10 ranked edges was not observed when considering edge number normalized FRC. In this case, the enriched species in obese individuals with IBD were *Candidatus Sulcia muelleri*, similar to healthy, obese individuals, making differentiation challenging. Finally, we measured the runtimes for all algorithms: FRC was 3-4 orders of magnitude slower than the fastest algorithms, but it was almost 7 times faster than ORC (Tab 10).

Table 8. Network metrices of a graph, derived from a microbial community. The graphs for healthy, and IBD conditions, and for lean and obese states, had identical metrices, since they were derived from the same microbial community, but with different node and edge weights, as explained in the methods.

Metric	
Number of nodes	154
Number of edges	23188
Density	0.9841

(a)



(b)

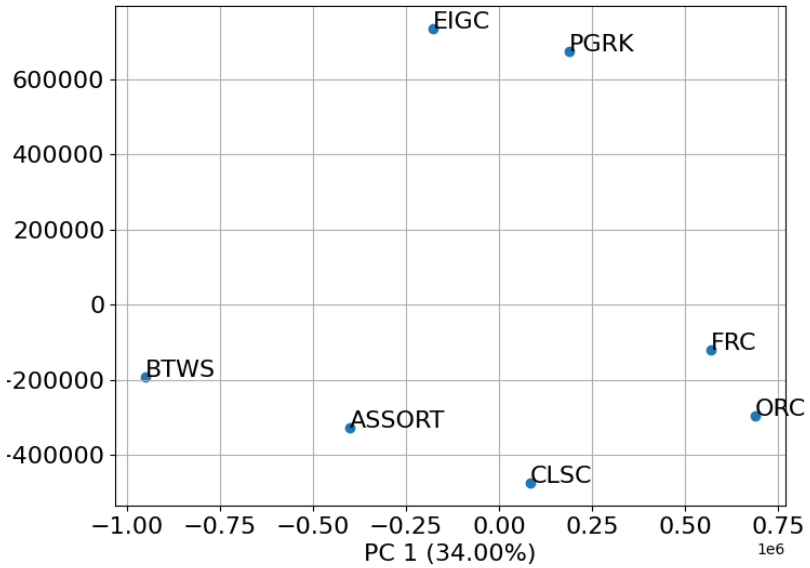
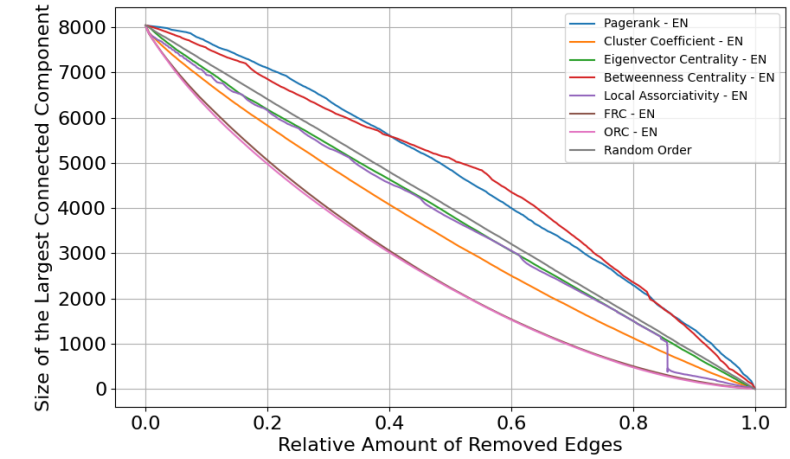


Fig 16. Comparison of seven edge ranking algorithms using a graph derived from a microbial community (a) Comparison of seven edge ranking algorithms based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS: edge betwenness, ASSORT: edge local assortiativity, PGRK: edge pagerank, EIGC: edge eigenvector centrality, CLSC: edge cluster coefficient, ORC: Ollivier-Ricci Curvature, FRC: Forman-Ricci Curvature.

(a)



(b)

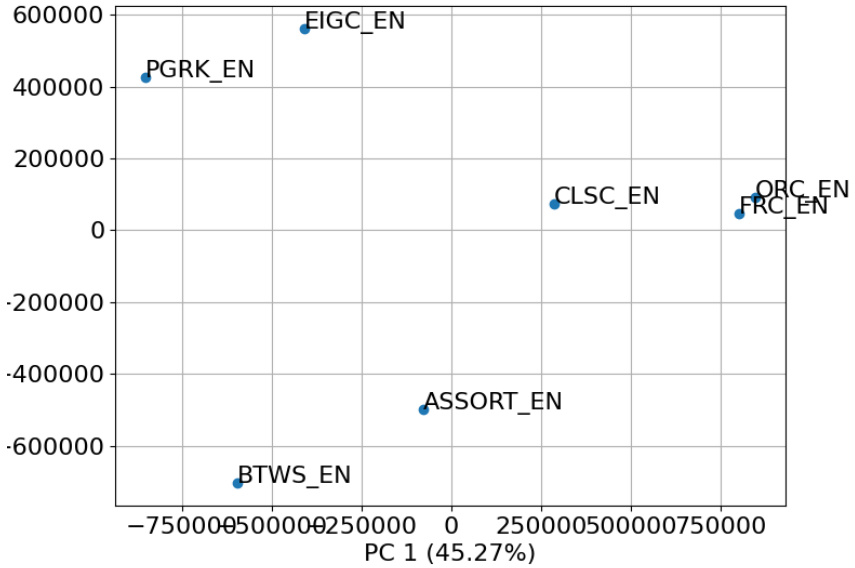


Fig 17. Comparison of seven edge ranking algorithms considering the normalization for edge number using a graph derived from a microbial community (a) Comparison of seven edge ranking algorithms with normalization for edge number based on their impact on the size of the weighted LCC upon the removal of the most important edges identified by each algorithm. Each algorithm, detailed in the Methods section, has been adapted for edge ranking with normalization for edge number as necessary. (b) PCA of edges assigned to ranks, according to each algorithm to display the similarity in edge-rankings between methods. BTWS_EN: edge betwenness, ASSORT_EN: edge local assortiativity, PGRK_EN: edge pagerank, EIGC_EN: edge eigenvector centrality, CLSC_EN: edge cluster coefficient, ORC_EN: Ollivier-Ricci Curvature, FRC_EN: Forman-Ricci Curvature, EN: normalized for edge number.

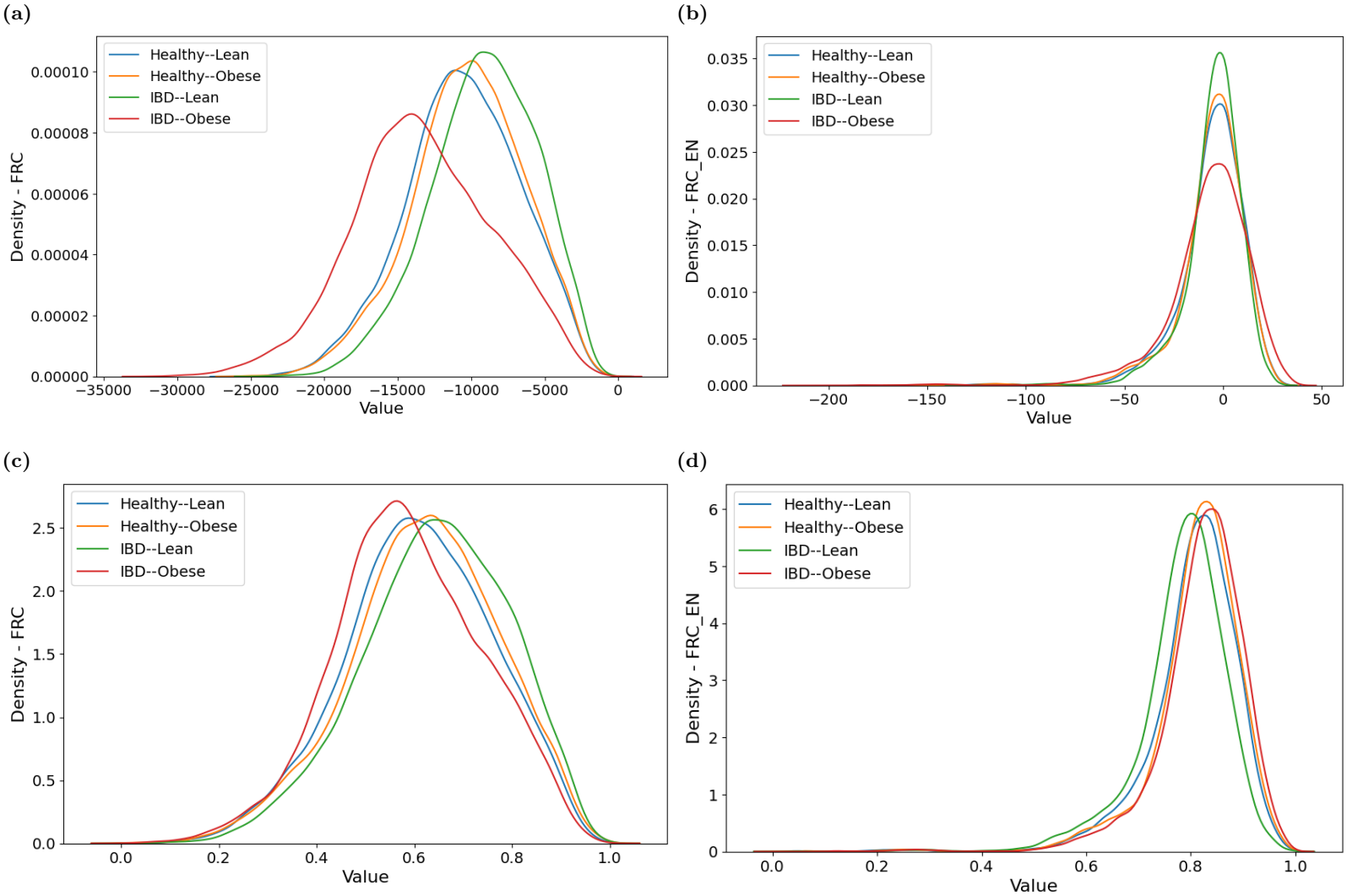


Fig 18. Distribution of FRCs, assigned to edges in the microbial community. (a) Distribution of four graphs, derived from microbial communities, (b) when considering edge number normalization, (c) when normalizing the distribution, and (d) when considering edge number normalization, and normalizing the distribution. The best separation of distributions was possible with the original values in figure (a).

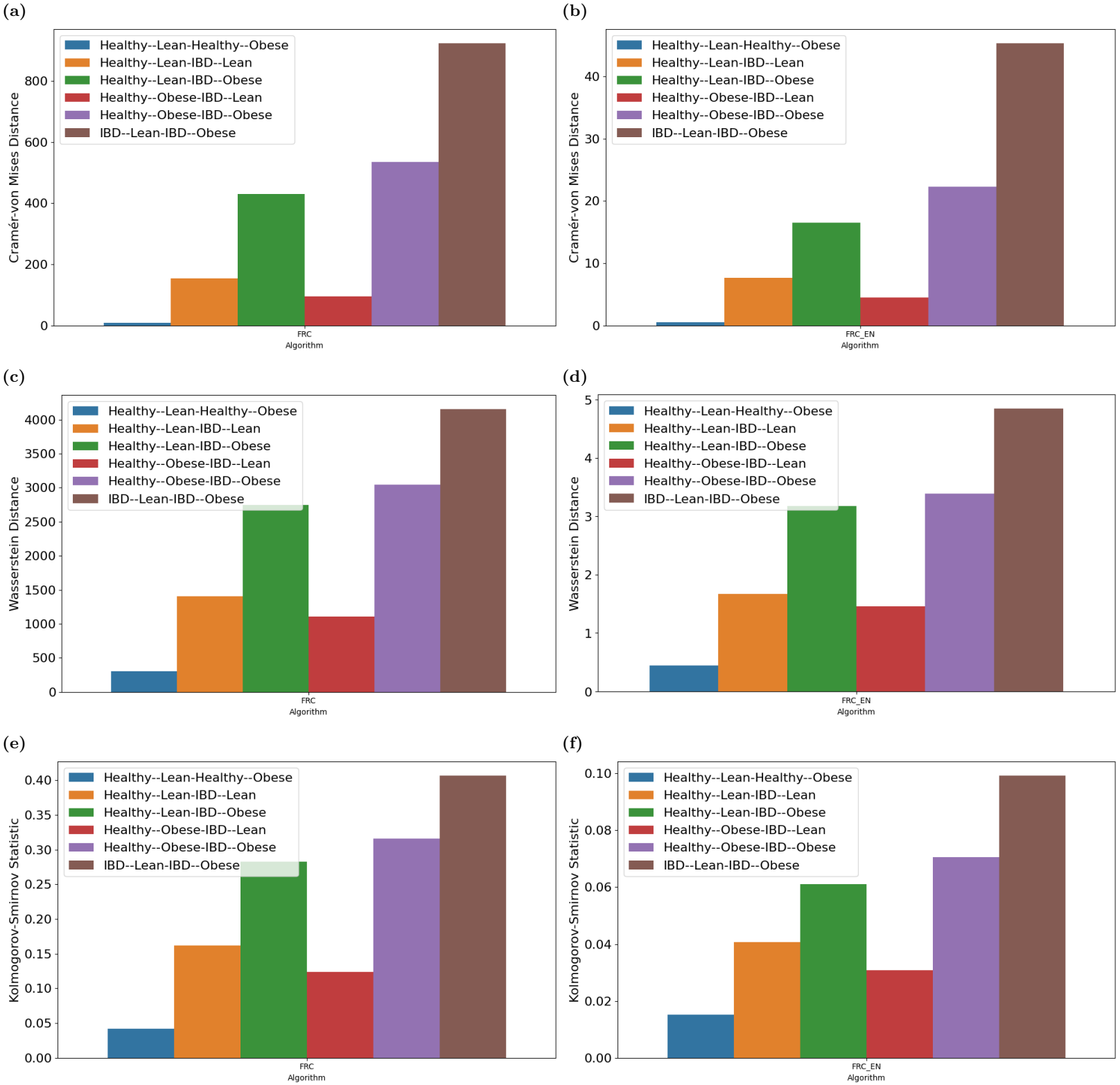


Fig 19. Evaluation of FRC and its capability to differentiate between networks Cramer-von Mises distance (a, b), Wasserstein distance (c, d), and Kolmogorov-Smirnow test (e, f) to identify which algorithm captures the difference of graphs, derived from microbial communities better, when considering original distributions (a, c, e), or normalized distributions (b, d, f).

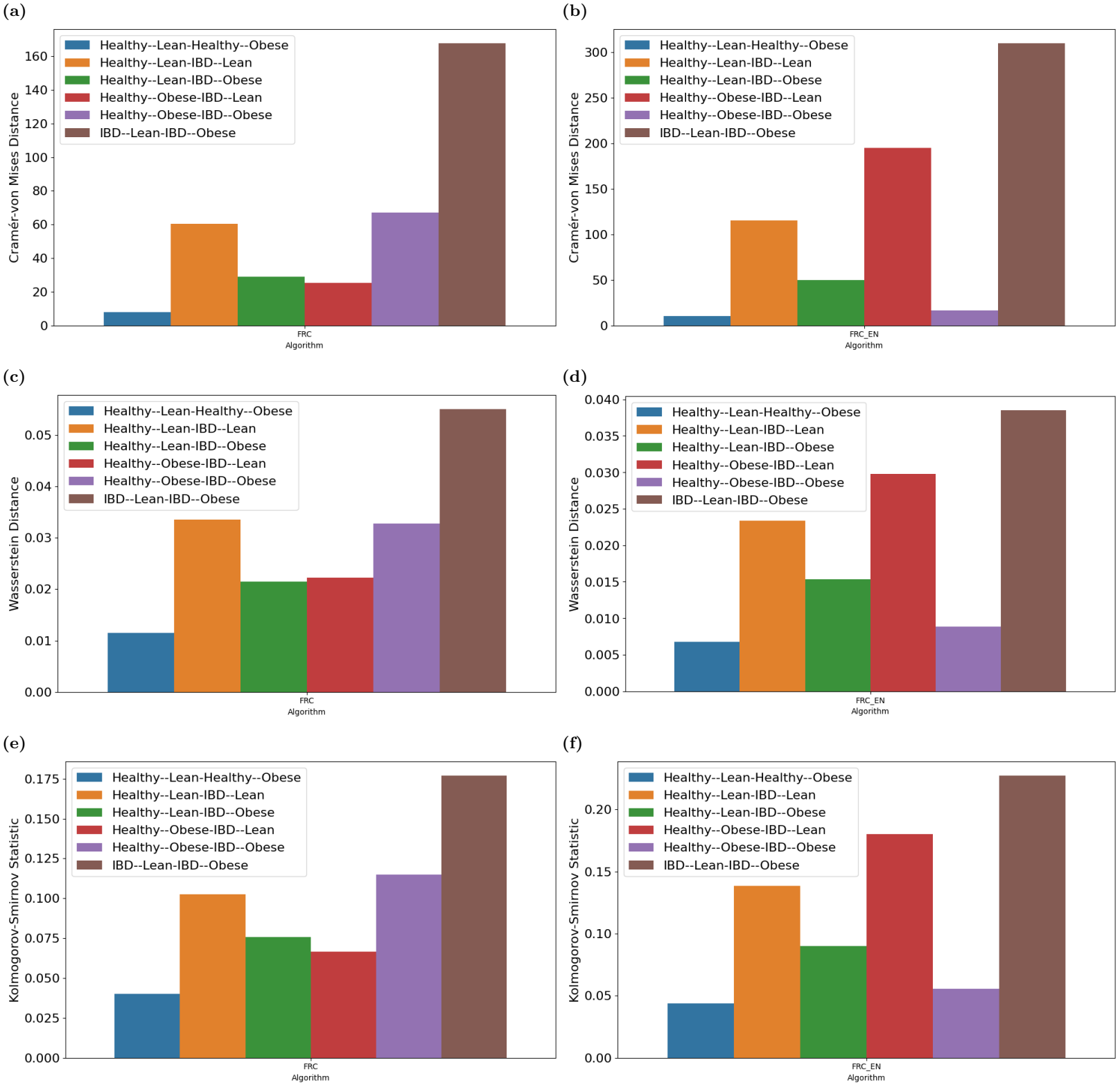


Fig 20. Evaluation of FRC and its capability to differentiate between networks, when considering edge number normalization normalization (EN) Cramer-von Mises distance (a, b), Wasserstein distance (c, d), and Kolmogorov-Smirnow test (e, f) to identify which algorithm captures the difference of graphs, derived from microbial communities better, when considering original distributions (a, c, e), or normalized distributions (b, d, f).

Table 9. Top 10 ranked edges obtained from FRC for every microbial network. Top-ranked edges from FRC for all networks were extracted considering the original implementation. While *Candidatus Sulcia muelleri* dominates the most important edges in healthy, and IBD–lean conditions, it only plays a minor role in the 10 top-ranked edges in the IBD–obese network. Also *Gordonibacter pamelaeae* is only present within the top 10 ranked edges in the IBD networks, which was confirmed by [95].

Condition	Species 1	Species 2
Healthy–Lean	Candidatus Sulcia muelleri	Bacteroides sp 9 1 42FAA
	Candidatus Sulcia muelleri	Bacteroides sp 4 3 47FAA
	Candidatus Sulcia muelleri	Bacteroides dorei DSM 17855
	Candidatus Sulcia muelleri	Bacteroides sp 2 2 4
	Candidatus Sulcia muelleri	Bacteroides vulgatus ATCC 8482
	Candidatus Sulcia muelleri	Bacteroides ovatus ATCC 8483
	Candidatus Sulcia muelleri	Bacteroides dorei 5 1 36 D4
	Candidatus Sulcia muelleri	Coprococcus comes ATCC 27758
	Candidatus Sulcia muelleri	Faecalibacterium prausnitzii SL3 3
	Collinsella aerofaciens ATCC 25986	Citrobacter koseri ATCC BAA 895
Healthy–Obese	Candidatus Sulcia muelleri	Bacteroides sp 9 1 42FAA
	Collinsella intestinalis DSM 13280	Citrobacter koseri ATCC BAA 895
	Candidatus Sulcia muelleri	Bacteroides sp 4 3 47FAA
	Candidatus Sulcia muelleri	Bacteroides sp 2 2 4
	Candidatus Sulcia muelleri	Bacteroides dorei DSM 17855
	Collinsella intestinalis DSM 13280	Enterobacter sp 638
	Candidatus Sulcia muelleri	Bacteroides vulgatus ATCC 8482
	Collinsella intestinalis DSM 13280	Cronobacter sakazakii ATCC BAA 894
	Collinsella intestinalis DSM 13280	Klebsiella pneumoniae 342
	Collinsella intestinalis DSM 13280	Salmonella enterica sv Heidelberg SL476 CVM30485
IBD–Lean	Candidatus Sulcia muelleri	Bacteroides dorei DSM 17855
	Candidatus Sulcia muelleri	Bacteroides vulgatus ATCC 8482
	Candidatus Sulcia muelleri	Bacteroides sp 4 3 47FAA
	Candidatus Sulcia muelleri	Bacteroides dorei 5 1 36 D4
	Candidatus Sulcia muelleri	Bacteroides sp 2 2 4
	Candidatus Sulcia muelleri	Bacteroides sp 9 1 42FAA
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides dorei DSM 17855
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides vulgatus ATCC 8482
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides sp 4 3 47FAA
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides dorei 5 1 36 D4
IBD–Obese	Collinsella intestinalis DSM 13280	Klebsiella pneumoniae 342
	Collinsella intestinalis DSM 13280	Escherichia fergusonii UMN026 ATCC 35469
	Collinsella intestinalis DSM 13280	Salmonella enterica sv Heidelberg SL476 CVM30485
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Klebsiella pneumoniae 342
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides sp 4 3 47FAA
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides dorei 5 1 36 D4
	Candidatus Sulcia muelleri	Bacteroides sp 4 3 47FAA
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides vulgatus ATCC 8482
	Candidatus Sulcia muelleri	Bacteroides dorei 5 1 36 D4
	Gordonibacter pamelaeae 7 10 1 bT DSM 19378	Bacteroides sp 2 1 7

Table 10. Runtimes of algorithms applied to a graph, derived from microbial communities. The runtime was measured for each algorithm, when using the original implementation (Original), and the augmented algorithm when normalizing for edge number normalization (EN). The fastest algorithm was edge pagerank, and edge eigenvector centrality, while the slowest algorithm were FRC and ORC.

Algorithm / Augmentation	Original	EN
Egde Betweenneess	2.01	2.06
Edge Pagerank	0.42	0.11
Edge Eigenvector Centrality	0.07	0.10
Edge Cluster Coefficient	64.79	65.35
Edge Local Assorciativity	4.02	4.07
Forman-Ricci-Curvature	595.458	605.21
Ollivier-Ricci-Curvature	4376.17	4325.76

Conclusion

In this study, we presented the Forman-Ricci curvature (FRC) applied to biological networks and evaluated its ability to differentiate metabolic and microbial graphs compared to other edge ranking algorithms, as well as its augmentation using edge number normalization. Biological graphs pose a challenge due to their inherent characteristics of directed edges and weights on both edges and nodes. The necessity of algorithms that can integrate all this information is critical, as existing algorithms tend to identify cofactor-associated hubs in metabolic graphs, for example. To address this issue, we applied the FRC to metabolic and microbial graphs and compared its performance with other algorithms. Additionally, we augmented the original FRC algorithm by normalizing edge terms by the incidence of the corresponding node. This approach provides more detailed insights into the network and prevents the identification of cofactors as hub edges.

The edge number normalization improved the separation of metabolic graphs, but the results were less favorable for microbial graphs. When using a weighted LCC approach, edge number normalization enhanced the results across all networks for FRC. In terms of the potential to separate FRC distributions, edge number normalization yielded better outcomes for metabolic graphs but not for microbial graphs. This discrepancy may be attributed to the network structures: metabolic graphs are less dense and more loosely connected, whereas microbial graphs are highly connected and dense, where the edge number has less impact. Furthermore, in metabolic graphs, only a few nodes are weighted with measured metabolites, while others are assigned placeholder numbers, which may influence algorithm performance. In contrast, microbial graphs have specific abundance values for each node, or nodes can be removed if absent from the network.

Although further studies are necessary on larger metabolic networks from mammalian GSMMs and on larger microbial networks, such as the most recent Agora2 [98], we have benchmarked the FRC for metabolic and microbial networks, laying the groundwork for future research.

References

1. Levy R, Borenstein E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. Proceedings of the National Academy of Sciences. 2013;110(31):12804–12809.

2. Barabasi AL, Oltvai ZN. Network biology: understanding the cell's functional organization. Nature reviews genetics. 2004;5(2):101–113.

3. Newman ME. The structure and function of complex networks. SIAM review. 2003;45(2):167–256.

4. Boccaletti S, Latora V, Moreno Y, Chavez M, Hwang DU. Complex networks: Structure and dynamics. Physics reports. 2006;424(4-5):175–308.

5. Chen SJ, Liao DL, Chen CH, Wang TY, Chen KC. Construction and analysis of protein-protein interaction network of heroin use disorder. Scientific reports. 2019;9(1):4980.

6. Safari-Alighiarloo N, Taghizadeh M, Rezaei-Tavirani M, Goliaei B, Peyvandi AA. Protein-protein interaction networks (PPI) and complex diseases. Gastroenterology and Hepatology from bed to bench. 2014;7(1):17.

7. Nabieva E, Jim K, Agarwal A, Chazelle B, Singh M. Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. Bioinformatics. 2005;21(suppl_1):i302–i310.

8. Koschützki D, Schreiber F. Centrality analysis methods for biological networks and their application to gene regulatory networks. Gene regulation and systems biology. 2008;2:GRSB–S702.

9. Beguerisse-Díaz M, Bosque G, Oyarzún D, Picó J, Barahona M. Flux-dependent graphs for metabolic networks. NPJ systems biology and applications. 2018;4(1):32.

10. Wagner A, Fell DA. The small world inside large metabolic networks. Proceedings of the Royal Society of London Series B: Biological Sciences. 2001;268(1478):1803–1810.

11. Yanashima R, Kitagawa N, Matsubara Y, Weatheritt R, Oka K, Kikuchi S, et al. Network features and pathway analyses of a signal transduction cascade. Frontiers in Neuroinformatics. 2009;3:360.

12. Matchado MS, Lauber M, Reitmeier S, Kacprowski T, Baumbach J, Haller D, et al. Network analysis methods for studying microbial communities: A mini review. Computational and structural biotechnology journal. 2021;19:2687–2698.

13. Faust K. Open challenges for microbial network construction and analysis. The ISME Journal. 2021;15(11):3111–3118.

14. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome biology. 2019;20:1–18.

15. Schilling CH, Palsson BØ. Assessment of the metabolic capabilities of Haemophilus influenzae Rd through a genome-scale pathway analysis. Journal of theoretical biology. 2000;203(3):249–283.

16. Han Y, Rangel AT, Pomraning KR, Kerkhoven EJ, Kim J. Advances in genome-scale metabolic models of industrially important fungi. Current Opinion in Biotechnology. 2023;84:103005.

17. Széliová D, Ruckerbauer DE, Galleguillos SN, Petersen LB, Natter K, Hanscho M, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. Metabolic engineering. 2020;61:288–300.

18. Brunner M, Kolb K, Keitel A, Stiefel F, Wucherpfennig T, Bechmann J, et al. Application of metabolic modeling for targeted optimization of high seeding density processes. Biotechnology and Bioengineering. 2021;118(5):1793–1804.

19. Richelle A, David B, Demaegd D, Dewerchin M, Kinet R, Morreale A, et al. Towards a widespread adoption of metabolic modeling tools in biopharmaceutical industry: a process systems biology engineering perspective. NPJ systems biology and applications. 2020;6(1):6.

20. Nilsson A, Nielsen J. Genome scale metabolic modeling of cancer. Metabolic engineering. 2017;43:103–112.

21. Lewis JE, Kemp ML. Integration of machine learning and genome-scale metabolic modeling identifies multi-omics biomarkers for radiation resistance. Nature Communications. 2021;12(1):2700.

22. Gelbach PE, Finley SD. Genome-scale modeling predicts metabolic differences between macrophage subtypes in colorectal cancer. Iscience. 2023;26(9).

23. Hsieh YE, Tandon K, Verbruggen H, Nikoloski Z. Comparative analysis of metabolic models of microbial communities reconstructed from automated tools and consensus approaches. npj Systems Biology and Applications. 2024;10(1):54.

24. Giordano N, Gaudin M, Trottier C, Delage E, Nef C, Bowler C, et al. Genome-scale community modelling reveals conserved metabolic cross-feedings in epipelagic bacterioplankton communities. Nature Communications. 2024;15(1):2721.

25. Robinson JL, Kocabaş P, Wang H, Cholley PE, Cook D, Nilsson A, et al. An atlas of human metabolism. Science signaling. 2020;13(624):eaaz1482.

26. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. Nature biotechnology. 2018;36(3):272–281.

27. Hefzi H, Ang KS, Hanscho M, Bordbar A, Ruckerbauer D, Lakshmanan M, et al. A consensus genome-scale reconstruction of Chinese hamster ovary cell metabolism. Cell systems. 2016;3(5):434–443.

28. Wang H, Robinson JL, Kocabas P, Gustafsson J, Anton M, Cholley PE, et al. Genome-scale metabolic network reconstruction of model animals as a platform for translational research. Proceedings of the National Academy of Sciences. 2021;118(30):e2102344118.

29. Strain B, Morrissey J, Antonakoudis A, Kontoravdi C. How reliable are Chinese hamster ovary (CHO) cell genome-scale metabolic models? Biotechnology and Bioengineering. 2023;120(9):2460–2478.

30. Vieira V, Ferreira J, Rocha M. A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale. PLoS computational biology. 2022;18(6):e1009294.

31. Li F, Chen Y, Qi Q, Wang Y, Yuan L, Huang M, et al. Improving recombinant protein production by yeast through genome-scale modeling using proteome constraints. Nature communications. 2022;13(1):2969.

32. Oftadeh O, Hatzimanikatis V. Genome-scale models of metabolism and expression predict the metabolic burden of recombinant protein expression. Metabolic Engineering. 2024;.

33. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature biotechnology. 2010;28(3):245–248.

34. Lewis NE, Hixson KK, Conrad TM, Lerman JA, Charusanti P, Polpitiya AD, et al. Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. Molecular systems biology. 2010;6(1):390.

35. Burgard AP, Vaidyaraman S, Maranas CD. Minimal reaction sets for Escherichia coli metabolism under different growth requirements and uptake environments. Biotechnology progress. 2001;17(5):791–797.

36. Bodeit O, Ben Samir I, Karr JR, Goelzer A, Liebermeister W. RBAtools: a programming interface for Resource Balance Analysis models. Bioinformatics Advances. 2023;3(1):vbad056.

37. Bulović A, Fischer S, Dinh M, Golib F, Liebermeister W, Poirier C, et al. Automated generation of bacterial resource allocation models. Metabolic engineering. 2019;55:12–22.

38. Bordel S, Agren R, Nielsen J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. PLoS computational biology. 2010;6(7):e1000859.

39. Cottret L, Jourdan F. Graph methods for the investigation of metabolic networks in parasitology. Parasitology. 2010;137(9):1393–1407.

40. Kreimer A, Borenstein E, Gophna U, Ruppin E. The evolution of modularity in bacterial metabolic networks. Proceedings of the National Academy of Sciences. 2008;105(19):6976–6981.

41. Croes D, Couche F, Wodak SJ, van Helden J. Inferring meaningful pathways in weighted metabolic networks. Journal of molecular biology. 2006;356(1):222–236.

42. Borzou P, Ghaisari J, Izadi I, Eshraghi Y, Gheisari Y. A novel strategy for dynamic modeling of genome-scale interaction networks. Bioinformatics. 2023;39(2):btad079.

43. Passi A, Tibocha-Bonilla JD, Kumar M, Tec-Campos D, Zengler K, Zuniga C. Genome-scale metabolic modeling enables in-depth understanding of big data. Metabolites. 2021;12(1):14.

44. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabási AL. The large-scale organization of metabolic networks. Nature. 2000;407(6804):651–654.

45. Zhu Q, Qin T, Jiang YY, Ji C, Kong DX, Ma BG, et al. Chemical basis of metabolic network organization. PLoS Computational Biology. 2011;7(10):e1002214.

46. Vitkup D, Kharchenko P, Wagner A. Influence of metabolic network structure and function on enzyme evolution. Genome biology. 2006;7:1–9.

47. Ma HW, Zhao XM, Yuan YJ, Zeng AP. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. Bioinformatics. 2004;20(12):1870–1876.

48. Yang P, Yu S, Cheng L, Ning K. Meta-network: optimized species-species network analysis for microbial communities. BMC genomics. 2019;20:143–151.

49. Loftus M, Hassouneh SAD, Yooseph S. Bacterial associations in the healthy human gut microbiome across populations. Scientific reports. 2021;11(1):2828.

50. Cardona C, Weisenhorn P, Henry C, Gilbert JA. Network-based metabolic analysis and microbial community modeling. Current Opinion in Microbiology. 2016;31:124–131.

51. Lam TJ, Ye Y. Meta-analysis of microbiome association networks reveal patterns of dysbiosis in diseased microbiomes. Scientific Reports. 2022;12(1):17482.

52. Jiang D, Armour CR, Hu C, Mei M, Tian C, Sharpton TJ, et al. Microbiome multi-omics network analysis: statistical considerations, limitations, and opportunities. Frontiers in genetics. 2019;10:995.

53. Blanchet FG, Cazelles K, Gravel D. Co-occurrence is not evidence of ecological interactions. Ecology Letters. 2020;23(7):1050–1063.

54. Guo B, Zhang L, Sun H, Gao M, Yu N, Zhang Q, et al. Microbial co-occurrence network topological properties link with reactor parameters and reveal importance of low-abundance genera. npj Biofilms and Microbiomes. 2022;8(1):3.

55. Estrada-Peña A, Cabezas-Cruz A, Pollet T, Vayssier-Taussat M, Cosson JF. High throughput sequencing and network analysis disentangle the microbial communities of ticks and hosts within and between ecosystems. Frontiers in Cellular and Infection Microbiology. 2018;8:236.

56. Zamkovaya T, Foster JS, de Crécy-Lagard V, Conesa A. A network approach to elucidate and prioritize microbial dark matter in microbial communities. The ISME journal. 2021;15(1):228–244.

57. Del Rio-Hortega L, Martín-Forés I, Castro I, De Miguel JM, Acosta-Gallo B. Network-based analysis reveals differences in plant assembly between the native and the invaded ranges. NeoBiota. 2022;72:157–181.

58. González AMM, Dalsgaard B, Olesen JM. Centrality measures and the importance of generalist species in pollination networks. Ecological complexity. 2010;7(1):36–43.

59. Leal W, Restrepo G, Stadler PF, Jost J. Forman–Ricci curvature for hypergraphs. Advances in Complex Systems. 2021;24(01):2150003.

60. Weber M, Saucan E, Jost J. Characterizing complex networks with Forman-Ricci curvature and associated geometric flows. arXiv preprint arXiv:160708654. 2016;.

61. Forman. Bochner's method for cell complexes and combinatorial Ricci curvature. Discrete & Computational Geometry. 2003;29:323–374.

62. Sreejith R, Jost J, Saucan E, Samal A. Systematic evaluation of a new combinatorial curvature for complex networks. Chaos, Solitons & Fractals. 2017;101:50–67.

63. Saucan E, Sreejith R, Vivek-Ananth R, Jost J, Samal A. Discrete Ricci curvatures for directed networks. Chaos, Solitons & Fractals. 2019;118:347–360.

64. Ollivier Y. Ricci curvature of Markov chains on metric spaces. Journal of Functional Analysis. 2009;256(3):810–864.

65. Samal A, Sreejith R, Gu J, Liu S, Saucan E, Jost J. Comparative analysis of two discretizations of Ricci curvature for complex networks. Scientific reports. 2018;8(1):8650.

66. Pouryahya M, Mathews J, Tannenbaum A. Comparing three notions of discrete Ricci curvature on biological networks. arXiv preprint arXiv:171202943. 2017;.

67. Fesser L, Iváñez SSdH, Devriendt K, Weber M, Lambiotte R. Augmentations of Forman's Ricci Curvature and their Applications in Community Detection. arXiv preprint arXiv:230606474. 2023;.

68. Murgas KA, Saucan E, Sandhu R. Quantifying cellular pluripotency and pathway robustness through Forman-Ricci curvature. In: Complex networks & their applications X: volume 2, proceedings of the tenth international conference on complex networks and their applications COMPLEX NETWORKS 2021 10. Springer; 2022. p. 616–628.

69. Tarjan R. Depth-first search and linear graph algorithms. SIAM journal on computing. 1972;1(2):146–160.

70. Bie L, Zhang M, Wang J, Fang M, Li L, Xu H, et al. Comparative analysis of transcriptomic response of Escherichia coli K-12 MG1655 to nine representative classes of antibiotics. Microbiology Spectrum. 2023;11(2):e00317–23.

71. Greenacre M, Groenen PJ, Hastie T, d'Enza AI, Markos A, Tuzhilina E. Principal component analysis. Nature Reviews Methods Primers. 2022;2(1):100.

72. Kim J, Cheong YE, Jung I, Kim KH. Metabolomic and transcriptomic analyses of Escherichia coli for efficient fermentation of L-fucose. Marine drugs. 2019;17(2):82.

73. Girvan M, Newman ME. Community structure in social and biological networks. Proceedings of the national academy of sciences. 2002;99(12):7821–7826.

74. Page L, Brin S, Motwani R, Winograd T. The PageRank citation ranking: Bringing order to the web. Stanford infolab; 1999.

75. Berman A, Plemmons RJ. Nonnegative matrices in the mathematical sciences. SIAM; 1994.

76. Saramäki J, Kivelä M, Onnela JP, Kaski K, Kertesz J. Generalizations of the clustering coefficient to weighted complex networks. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics. 2007;75(2):027105.

77. Thedchanamoorthy G, Piraveenan M, Kasthuriratna D, Senanayake U. Node assortativity in complex networks: An alternative approach. Procedia Computer Science. 2014;29:2449–2461.

78. Kramer O, Kramer O. Scikit-learn. Machine learning for evolution strategies. 2016; p. 45–53.

79. Anderson TW. On the distribution of the two-sample Cramer-von Mises criterion. The Annals of Mathematical Statistics. 1962; p. 1148–1159.

80. Piccoli B, Rossi F. On properties of the generalized Wasserstein distance. Archive for Rational Mechanics and Analysis. 2016;222:1339–1365.

81. Massey Jr FJ. The Kolmogorov-Smirnov test for goodness of fit. Journal of the American statistical Association. 1951;46(253):68–78.

82. Hagberg A, Swart PJ, Schult DA. Exploring network structure, dynamics, and function using NetworkX. Los Alamos National Laboratory (LANL), Los Alamos, NM (United States); 2008.

83. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: constraints-based reconstruction and analysis for python. BMC systems biology. 2013;7:1–6.

84. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. i ML1515, a knowledgebase that computes Escherichia coli traits. Nature biotechnology. 2017;35(10):904–908.

85. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic acids research. 2017;45(D1):D353–D361.

86. Li G, Zhu F, Wei P, Xue H, Chen N, Lu B, et al. Metabolic engineering of Escherichia coli for hyperoside biosynthesis. Microorganisms. 2022;10(3):628.

87. Buckstein MH, He J, Rubin H. Characterization of nucleotide pools as a function of physiological state in Escherichia coli. Journal of bacteriology. 2008;190(2):718–726.

88. Khakoo NS, Ioannou S, Khakoo NS, Vedantam S, Pearlman M. Impact of obesity on inflammatory bowel disease. Current gastroenterology reports. 2022;24(1):26–36.

89. Johnson AM, Loftus EV. Impact of obesity on the management of inflammatory bowel disease. Gastroenterology & Hepatology. 2020;16(7):350.

90. Szilagyi A. Relationship (s) between obesity and inflammatory bowel diseases: possible intertwined pathogenic mechanisms. Clinical Journal of Gastroenterology. 2020;13(2):139–152.

91. Islam MR, Arthur S, Haynes J, Butts MR, Nepal N, Sundaram U. The role of gut microbiota and metabolites in obesity-associated chronic gastrointestinal disorders. Nutrients. 2022;14(3):624.

92. Gomez-Arango LF, Barrett HL, Wilkinson SA, Callaway LK, McIntyre HD, Morrison M, et al. Low dietary fiber intake increases Collinsella abundance in the gut microbiota of overweight and obese pregnant women. Gut microbes. 2018;9(3):189–201.

93. Ayelén R, Pablo A, Sofía Q, Jimena C, Renata S, Carolina C, et al. New insights in Ulcerative Colitis Associated Gut Microbiota in South American Population: Akkermansia and Collinsella, two distinctive genera found in Argentine subjects. medRxiv. 2020; p. 2020–07.

94. Dahal RH, Kim S, Kim YK, Kim ES, Kim J. Insight into gut dysbiosis of patients with inflammatory bowel disease and ischemic colitis. Frontiers in Microbiology. 2023;14:1174832.

95. Wurdemann D, Tindall BJ, Pukall R, Lunsdorf H, Strompl C, Namuth T, et al. Gordonibacter pamelaeae gen. nov., sp. nov., a new member of the Coriobacteriaceae isolated from a patient with Crohn's disease, and reclassification of Eggerthella hongkongensis Lau et al. 2006 as Paraeggerthella hongkongensis gen. nov., comb. nov. International journal of systematic and evolutionary microbiology. 2009;59(6):1405–1415.

96. Opstelten JL, Plassais J, van Mil SW, Achouri E, Pichaud M, Siersema PD, et al. Gut microbial diversity is reduced in smokers with Crohn's disease. Inflammatory bowel diseases. 2016;22(9):2070–2077.

97. Liu S, Zhao W, Lan P, Mou X. The microbiome in inflammatory bowel diseases: from pathogenesis to therapy. Protein & cell. 2021;12(5):331–345.

98. Heinken A, Acharya G, Ravcheev DA, Hertel J, Nyga M, Okpala OE, et al. AGORA2: Large scale reconstruction of the microbiome highlights wide-spread drug-metabolising capacities. BioRxiv. 2020; p. 2020–11.