Dicle University
Journal of Engineering
https://dergipark.org.tr/tr/pub/dumf
duje.dicle.edu.tr

Research Article

Centrality with Entropy in Hypergraphs Based on Similarity Measures

İhsan TUĞAL[1*], Zeydin PALA[2]

[1] Muş Alparslan University, Software Engineering Department, i.tugal@alparslan.edu.tr, Orcid No: 0000-0003-1898-9438
[2] Muş Alparslan University, Software Engineering Department, z.pala@alparslan.edu.tr, Orcid No 0000-0002-2642-7788

ARTICLE INFO

Article history:

Received 24 January 2023
Received in revised form 5 July 2023
Accepted 6 July 2023
Available online 30 September 2023

Keywords:

centrality, hypergraphs, entropy, similarity

Doi: 10.24012/dumf.1241450

* Corresponding author

ABSTRACT

Hypergraphs and simplicial complexes can be used to model higher-order interactions. Graphs are limited to model and describe pairwise interactions. In this study, the issue of centrality in hypergraphs was studied. We introduce centrality measures based on the entropy of nodes and hyperedges in the hypergraphs. Until now, a lot of measures from various perspectives have been proposed to identify influential nodes, yet non provides a complete solution to the centrality problem. Because there are different perspectives on centrality. It is important to try different models to reach a solution in centrality problems. Entropy, which is a measure of uncertainty, is a guide in centrality measurements. It can produce ideal solutions for centrality. In complex systems, the entropy can be measured by different methods. In this study, the entropy calculation was made according to the union, intersection, and jaccard similarity values for nodes. The way that similarity is measured indicates the type of centrality. Local centralities were detected more precisely when the degree and union similarity values were used. The intersection and jaccard similarities showed us the global centralities. Traditional methods of centrality were also compared with the results of the proposed method. The accuracy of the method was tested with different hypergraph datasets. It has been shown that we can produce efficient results with different similarity parameters according to our wishes in hypergraphs.

Introduction

Graphs are very powerful modelling tools and widely used to represent data and systems in a connected world, yet representing systems with pairwise interactions via graphs is a challenge where describing group interactions explicitly is not possible [1]. However, real-world systems, such as collaboration networks, gene interaction networks, computer networks, social networks involve the interaction of more than two units, dependencies between more than two variables, or properties of collections of more than two objects [2]. For example, co-authorship publication data, collaboration datasets, chemical processes, neural systems seem to display group interactions. For such representative structures, hypergraphs and simplicial complexes are suitable for modeling, analyzing, and expression of complex systems [3]. Hypergraph theory was introduced by Berge [8]. It has a strong theory for solving real world problems. It can be used to model complex, group structured relational data. We can say that hypergraphs offer better solutions for modeling group interactions. Unlike a graph, the idea of a hypergraph is to widen the edges to connect more than two nodes as seen in Figure 1 [4]. Hypergraphs have identities related to a number of other mathematical structures important in data science, including finite topologies, simple complexes, and Sperner systems. A suitable structure enables the use of a wider range of mathematical methods. Some interactions, when simplified into pairwise relations and modeled using a graph, can exhibit a very different nature than when their actual complexity is modeled using a hypergraph [5].

We need to learn more information about the structure of systems. The similarity of system elements provides information about the structure. In this study, similarity of nodes or hyperedges was measured by using intersection and union which is set operations. Jaccard similarities of nodes and hyperedges were obtained. For centrality, these values were used for entropy calculations.

The entropy of node explains which nodes play an important role [6]. Entropy, which measures variation, amount of information, uncertainty, offers a convenient way of centrality. The entropy value of each node or hyperedge also provides a value for centrality. Parameters determine whether centrality is a local or global [7]. We know that the degree is a local measure. We will see

whether the centrality measures with union, intersection and jaccard values will give local or global results.

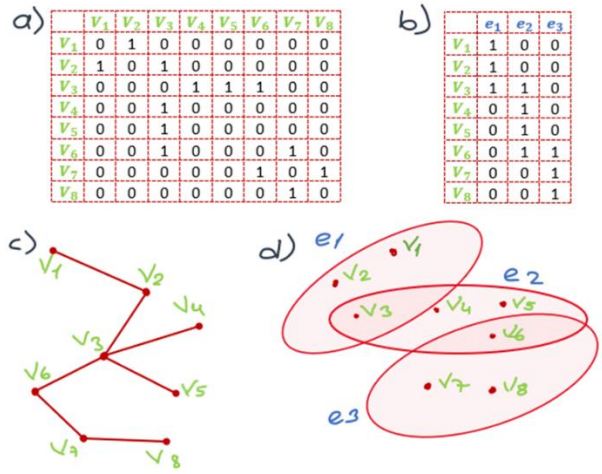


Figure 1. Representation of graph and hypergraph.

In this study, a method was proposed to identify influential nodes or hyperedges within the hypergraphs. The importance of the subject studied has been tried to be emphasized with the questions given below. What advantages do hypergraphs offer us? When should it be preferred to use hypergraphs instead of graphs? How to solve centrality problems with hypergraphs? The role of hypergraph models in solving various problems in higher-order systems was investigated.

The contribution of this study to the literature can be listed as follows:

1. A new method is proposed for the identification of central nodes and hyperedges in hypergraphs.

2. It has been shown that entropy is an efficient method for centrality calculations in hypergraphs.

3. It has been shown that node and hyperedge entropies can be measured using similarity values over neighborhood.

4. A method has emerged where hypergraph structures can be solved with same role for nodes and hyperedges.

5. Local and global centrality measures can be made by our method.

This paper is structured as follows: In section 2, the literature was reviewed. Section 3 describes the basic concepts that used in the proposed method. In Section 4, the method is explained with an example application. Section 5 presents experimental results and analyzes with 3 datasets, and Section 6 provides concluding remarks.

Preliminaries

In the study, Feng et al. found that hypergraph betweenness centrality is a superior method for identifying important genes. Hypergraph models were used to identify critical genes of biological networks. A comparison was made between the graphs and hypergraphs centrality measures. It was aimed to find critical genes effective in viral infection from genomic expression data sets. Networks extracted from correlation or mutual information [8].

Zhou et al. generalized the powerful spectral clustering methodology working on undirected graphs to hypergraphs. They developed algorithms for transductive classification and hypergraph embedding based on the spectral hypergraph clustering approach. The advantages of hypergraphs over graphs were demonstrated with applications [9].

In Ref. [10], it shows how the concept of network centrality can be adapted to supra-binary networks. The use of the technique was demonstrated by data on the attacks of inhabitants of the Caribbean islands on Spanish settlements in the period 1509-1700.

Hypergraph convolution and hypergraph attention which are two end-to-end trainable operators were introduced to the family of graph neural networks to efficiently learn deep embeddings on higher order graph-structured data. With two operators, a graph neural network is extended to a more flexible model and applied to various applications where non-binary relationships are observed. Extensive experimental results with semi-supervised node classification demonstrated the effectiveness of hypergraph convolution and hypergraph attention [11].

Ramadan et al used hypergraphs to model the yeast proteome, where proteins are nodes and complexes are hyperedges. To define the core proteome, they applied an algorithm that finds tightly connected nodes [12].

Zhou and Nakhleh examined the claim that metabolic networks are hierarchical and small-world networks. The reason for this claim was the graph model, Zhou and Nakleh modeled the metabolic networks of E. coli as hypergraphs rather than graphs, showing that the claimed hierarchy and scaling properties were not supported. This means: When biological interactions are simplified into binary relationships and modeled using a graph, they can exhibit a very different structure than when their actual state is modeled using a hypergraph [5].

Multi-morbidity refers to a health condition of having two or more concurrent chronic conditions. In the study [13], the centrality of the diseases according to the multiple status was measured using the hypergraph structure. Compared with the graph structure. It has been shown that hypergraphs can be used as a valuable tool for analyzing poorly understood relationships.

In the study [14], eigenvector centrality was applied to uniform hypergraphs. When the obtained results were examined, it was shown that hypergraphs could be

analyzed, and different information could be revealed about real-world data.

In the study[15], the authors argue that existing methods, which rely on the pairwise correlations between stocks, are not sufficient. They propose a new method that uses hypergraphs to model the relationships between stocks. The authors propose a new model called the hypergraph tri-attention network (HGTAN). HGTAN is a hierarchical attention model that can learn the importance of nodes, hyperedges, and hypergraphs during the information propagation among stocks. This allows HGTAN to fully exploit the potential synergies between stock movements.

The study [16] proposes a new method for learning brain functional connectome (FC) features using a dynamic weighted hypergraph convolutional network (dwHGCN). The dwHGCN is able to learn the importance of hyperedges and assign larger weights to hyperedges with higher discriminative power. This weighting strategy also improves the interpretability of the model by identifying the highly active interactions among ROIs shared by a common hyperedge.

Compared to the previous studies, here, we proposed new, effective, and accurate ways to measure centrality in hypergraphs. The similarities of the nodes and hyperedges were measured in the hypergraphs. Their entropies were found with these values and their centrality was shown.

Methodology

This section provides the necessary technical background, definitions and information for this work. The definitions of these methods and metrics are as follows.

Hypergraphs

A hypergraph is a generalization of a graph in which an edge can contain any number of nodes. Formally, a hypergraph H can be defined as $H = (V, E)$, where V is the set of nodes and E is the set of non-empty subsets of V called hyperedges or edges. Therefore, E is a subset of $P(V)$, where $P(V)$ is the power set of V. In graphs, edges connect pair of nodes, while hyperedges are sets of nodes and thus can contain many nodes depending on the situation.

An $H = (V, E)$ undirected hypergraph has the sets $V = \{v_1, v_2, \dots, v_n\}$ and $E = \{e_1, e_2, \dots, e_m\}$. The union of the e_i elements give the set V. Then H has an incidence matrix with $n \times m$ dimension. The incidence matrix ($I(H)$) is a matrix that shows the relationship between two object classes. If the first class is V and the second is E, the matrix has one row for each element of V and one column for each element of E. For any value of i and j in the I_{ij} incidence matrix, if $I_{ij} = 1$ then $v_i \in e_j$, otherwise $I_{ij} = 0$. These values can be taken as weighted values.

$$I_{ij} = \begin{cases} 1 & if\ v_i \in e_j \\ 0 & if\ v_i \notin e_j \end{cases} \quad (1)$$

Some data structures can be given as an example to understand the hypergraphs. For example, a hypergraph can be created where each node represents an author and each hyperedge represents an article written by several co-authors. An article with k authors has a hyperedge containing k nodes. This way, authors with closer collaboration are connected by more hyperedges. Considering a human brain network, if brain regions are defined as nodes, the neurological interaction of these brain regions can be expressed as a hyperedge. If proteins are defined as nodes and protein complexes can be thought of as hyperedges.

The hypergraphs provide a rich modeling framework, algorithms will necessarily be problem-specific and differ in complexity from algorithms used for graphs [17].

Similarity

The similarity rates of the nodes give us various information about the network. Similarity metrics are important for recommendation systems, link prediction etc. Similarity values used in this study were measured over neighborhoods. Two nodes are neighbors if there is an edge connecting them. If they have large overlap between their neighboring set of nodes, nodes are considered similar [18]. The common neighbor adopts the idea that two strangers with mutual friends are more likely to meet than those without mutual friends. Nodes with many common neighbors have high similarity. Most similarity-based algorithms use common neighbor information. It is widely used in the predictions of whom nodes can connect with in the future [19][20].

Let x and y be nodes. Let $\Gamma(x)$ be the set of nodes connected to x. Let $\Gamma(y)$ be the set of nodes connected by y. Accordingly, the common neighbor value is the intersection of the two sets as,

$$CN_{xy} = S_{xy} = |\Gamma(x) \cap \Gamma(y)| \quad (2)$$

The sum of neighbors is the union of the sets of neighbors of x and y as,

$$SN_{xy} = S_{xy} = |\Gamma(x) \cup \Gamma(y)| \quad (3)$$

Jaccard similarity is the number of common neighbors of two nodes divided by the number of union sets of neighbors. If two nodes have the same neighbors, the

jaccard similarity value is 1. Conversely, if they have no common neighbors, their similarity is 0.

$$Jaccard_{xy} = S_{xy} = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (4)$$

Entropy

The traditional measure of uncertainty is entropy, commonly referred to as the measure of Shannon information [21]. The entropy value tells the uncertainty of a system or, conversely, the amount of information it contains. It is possible to measure and interpret the network complexity in this way. It is a probabilistic calculation method [22]–[24]. If there is a probability of n states occurring in a system, and the probability of each state occurring is represented by p, the entropy of the system is calculated as

$$Entropy = -\sum_{i=1}^{n} p_i log_2 p_i \quad (5)$$

When one of the probabilities is 1 and the others are 0, its value is zero and the entropy is minimal. If all probability values are the same and $p_1 = p_2 = \cdots = p_n = 1/n$, then entropy is also maximum.

While calculating entropy, node or hyperedge degrees in the incidence matrix (I_{ij}) is used in the equations below. The link weight of $node\ i$ and $hyperedge\ j$ is represented by the value w_{ij}. It must be noted that the node and the hyperedge can switch roles.

$$d(v_i) = \sum_{j=1}^{m} w_{ij},\ d(e_i) = \sum_{j=1}^{n} w_{ij} \quad (6)$$

$$p(v_{ij}) = \frac{w_{ij}}{\sum_{j=1}^{m} w_{ij}},\ p(e_{ij}) = \frac{w_{ij}}{\sum_{j=1}^{n} w_{ij}} \quad (7)$$

$$Entropy(\boldsymbol{v_i}) = -\sum_{j=1}^{m} p(v_{ij}) log_2 p(v_{ij}),\quad Entropy(\boldsymbol{e_i}) = -\sum_{j=1}^{n} p(e_{ij}) log_2 p(e_{ij}) \quad (8)$$

How to measure the entropy has attracted much attention [25]. The entropy calculation was made according to the union, intersection and jaccard similarity values. It transforms from incidence matrix to weighted adjacency matrix (A) that is, similarity matrix (S) in a sense while performing union, intersection and jaccard calculations. The incidence matrix of measure nxm becomes nxn for nodes and mxm for hyperedges. The method is better understood if the sample application in Section 4 is examined. Let S_{ij} be our jaccard, intersection and union similarity matrix. Let it be nxn in size for nodes. Probability values for entropy are calculated as

$$p(v_{ij}) = \frac{s_{ij}}{\sum_{j=1}^{n} s_{ij}},\quad p(e_{ij}) = \frac{s_{ij}}{\sum_{j=1}^{m} s_{ij}} \quad (9)$$

Centrality

Centrality measures are used to identify the most important nodes of a network. While expressing centrality, we can describe it as influential nodes in a social network, nodes that accelerate the spread of information, points we call hubs in a transportation network, nodes that strengthen the network or key elements that direct the operation of the system. Generally, centrality measures are used in networks to estimate the potential monitoring and control capabilities that nodes may have over the communication flowing in the network [26]. In the dynamic behavior of the network system, the effect of central nodes in the organizational structure is more [27], [28].

Centrality is one of the most studied topics in network analysis. Many methods for centrality have been proposed in the literature to analyze the internal topology of a particular network and to find active nodes. Degree [29], closeness [30], betweenness [30], eigenvector [31] and pagerank [32] are the most commonly used measures of centrality [33]–[37].

Degree centrality allows us to evaluate based on the number of neighbors. It doesn't offer much information about the network. It is restricted to the node.

Betweenness centrality looks at how many times the shortest path between any pair of nodes in the network passes over a node and selects critical nodes accordingly. The critical node indicates how well it mediates the communication of other nodes.

Closeness centrality is a measure in a network that is calculated as the sum of the length of the shortest paths between the node and all other nodes in the network. The central node is the closer to the other nodes. These nodes have the shortest distances to all other nodes and can spread information very efficiently through a network.

This eigenvector centrality is based on the value of the immediate neighbours of each node. The importance of a node is related to how important its neighbors are. It is based on the eigenvalue. The value of a node depends on the value of the nodes linked to it. This measure is not preferred in large networks [38].

The purpose of pagerank is to identify nodes that have many connections to other important nodes. It indicates the probability that the node will be visited in a random walk. It works better for directional networks. It is similar to eigenvector centrality.

Entropy is used in centrality measurements. If we want to measure the effect of a node on the network, entropy values can guide us in this regard. Entropy centrality allows for state-of-the-art centrality in centrality analyses, thus distinguishing network nodes locally central and globally central. It generates the centrality result according to the selected link parameters.

Method with Sample Application

In our study, analysis was performed on data suitable for the hypergraphs. Most tabular real-world data are in this format. The proposed method was tried to be shown on the sample data set. The synthetic data with 4 nodes and 3 hyperedges used in this study are given in Table 1 and visualization in Figure 2.

Table 1. Incidence matrix of the sample dataset

	e_0	e_1	e_2	Node degree
v_0	1	1	1	**3**
v_1	1	0	1	**2**
v_2	1	1	0	**2**
v_3	0	0	1	**1**
v_4	1	1	0	**2**
Hyperedge degree	**4**	**3**	**3**	**Sum=10**

The nodes degree or hyperedges degree are calculated as follows.

$$Node_{degree} = d(v_i) = \sum_{j=1}^{m} w_{ij} \quad (10)$$

$$Hyperedge_{degree} = d(e_i) = \sum_{j=1}^{n} w_{ij} \quad (11)$$

In hypergraphs, the roles can be changed for hyperedges and nodes in the process steps. In the example, operations for nodes can be done for hyperedges. We denote the transpose of Incidence matrix I by I^T.

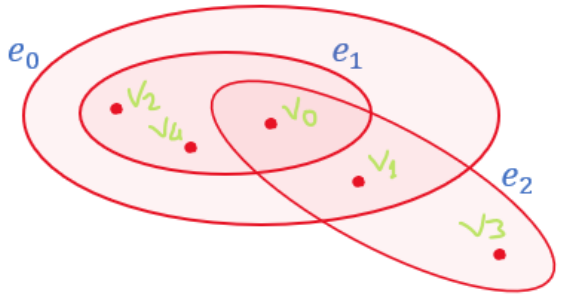


Figure 2. Hypergraph representation of the sample dataset

The intersection and union similarity values were found from the incidence matrix as in Table 2. In addition, one can state that a similarity matrix of nodes and hyperedges was created by intersection or union values. In a sense, hypergraphs were converted into graphs mode.

The intersection values of the nodes were calculated according to how many common hyperedges they were members of. For example, if the value of $intersect(v_0, v_1)$ is 2, that means there are 2 hyperedges where both nodes are elements. The union values of the nodes were calculated according to the total number of hyperedges that these nodes are elements. It suffices for one or both of the nodes to be elements of the hyperedge, for the hyperedge's interior. For example, when $union(v_0, v_1)$ is 3, it indicates that there are 3 hyperedges where either or both two nodes are members.

Table 2. Intersection and union similarity matrices

Nodes intersection						**Nodes union**					
	v_0	v_1	v_2	v_3	v_4		v_0	v_1	v_2	v_3	v_4
v_0	0	2	2	1	2	v_0	0	3	3	3	3
v_1	2	0	1	1	1	v_1	3	0	3	2	3
v_2	2	1	0	0	2	v_2	3	3	0	3	2
v_3	1	1	0	0	0	v_3	3	2	3	0	3
v_4	2	1	2	0	0	v_4	3	3	2	3	0

Jaccard similarity values in Table 3 are also calculated from intersection and union similarity values with Eq. 4. From here, the effect of the nodes and hyperedges in the system was identified by the help of entropy. Degree, intersection, union and jaccard similarity values were used while calculating entropy. Accuracy may vary on status as there are different perspectives on centrality calculations. The nodes or hyperedges with a high entropy value are the effective ones.

Table 3. Jaccard similarity matrix

	v_0	v_1	v_2	v_3	v_4
v_0	0	0,67	0,67	0,33	0,67
v_1	0,67	0	0,33	0,5	0,33
v_2	0,67	0,33	0	0	1
v_3	0,33	0,5	0	0	0
v_4	0,67	0,33	1	0	0

When examined Table 4, it shows the order of centrality with jaccard as $v0 > v1 > v2 = v4 > v3$. As it can be seen in the Figure 2, those with the same degrees received equal value in degree centrality. Sorting according to the results obtained with intersection and jaccard was similar. For jaccard and intersection on hyperedges, there is an ordering as $e0 > e2 > e1$. When we look at the entropy values calculated with degrees at the hyperedges, the degrees e1 and e2 are the same. An anomaly has occurred in the union entropy. Jaccard and intersection again showed similar behavior.

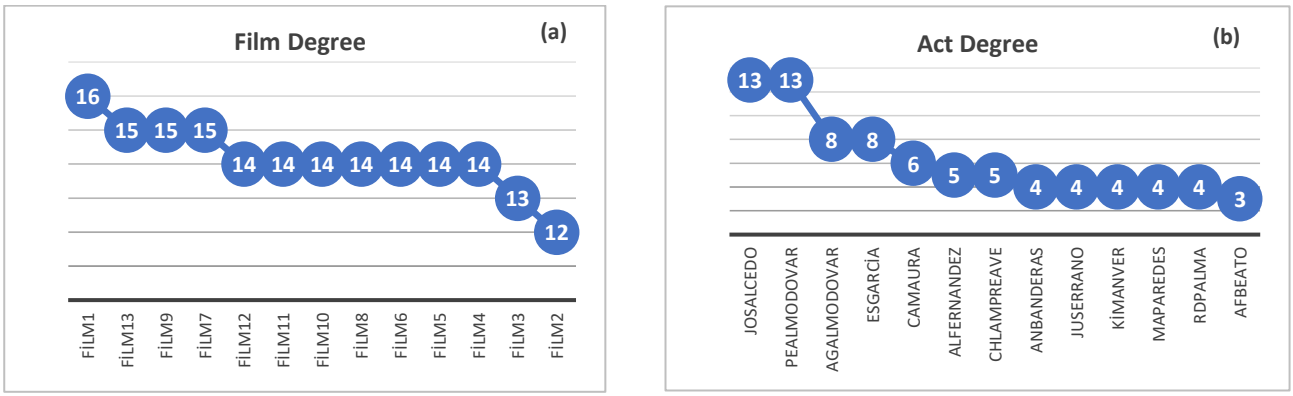
Table 4. Entropy values with different parameters

Nodes entropy								
	e_degree		e_intersection		e_union		e_jaccard	
v_0	**1,58**	v_0	**1.95**	v_0	**2**	v_0	**1,95**
v_1	1	v_1	1.92	v_1	1,98	v_1	1,94
v_2	1	v_2	1.52	v_2	1,98	v_2	1,46
v_3	0	v_3	1	v_3	1,98	v_3	0,97
v_4	1	v_4	1.52	v_4	1,98	v_4	1,46

Hyperedges entropy								
	e_degree		e_intersection		e_union		e_jaccard	
e_0	**2**	e_0	**0,97**	e_0	0,99	e_0	**0,93**
e_1	1,58	e_1	0,81	e_1	0,99	e_1	0,74
e_2	1,58	e_2	0,92	e_2	**1**	e_2	0,92

Experimental Results

The proposed model for centrality was applied to three different datasets. Two of them are Almoincidence and Bcancer datasets in hypergraph structure from Ucinet [39]. The other is a drug-target interaction network that contains information about which genes (i.e., the proteins encoded by the genes) are targeted by drugs on the US market [40]. Results show the effectiveness of our proposed method. Since our aim is to calculate centrality, examining degree centrality values, which is the most basic centrality calculation, can be supportive to evaluate the results. These values can be seen in the Figure 3. Almoincidence dataset contains weighted values. In this dataset, the results are calculated and evaluated using the weighted values.



Almoincidence dataset

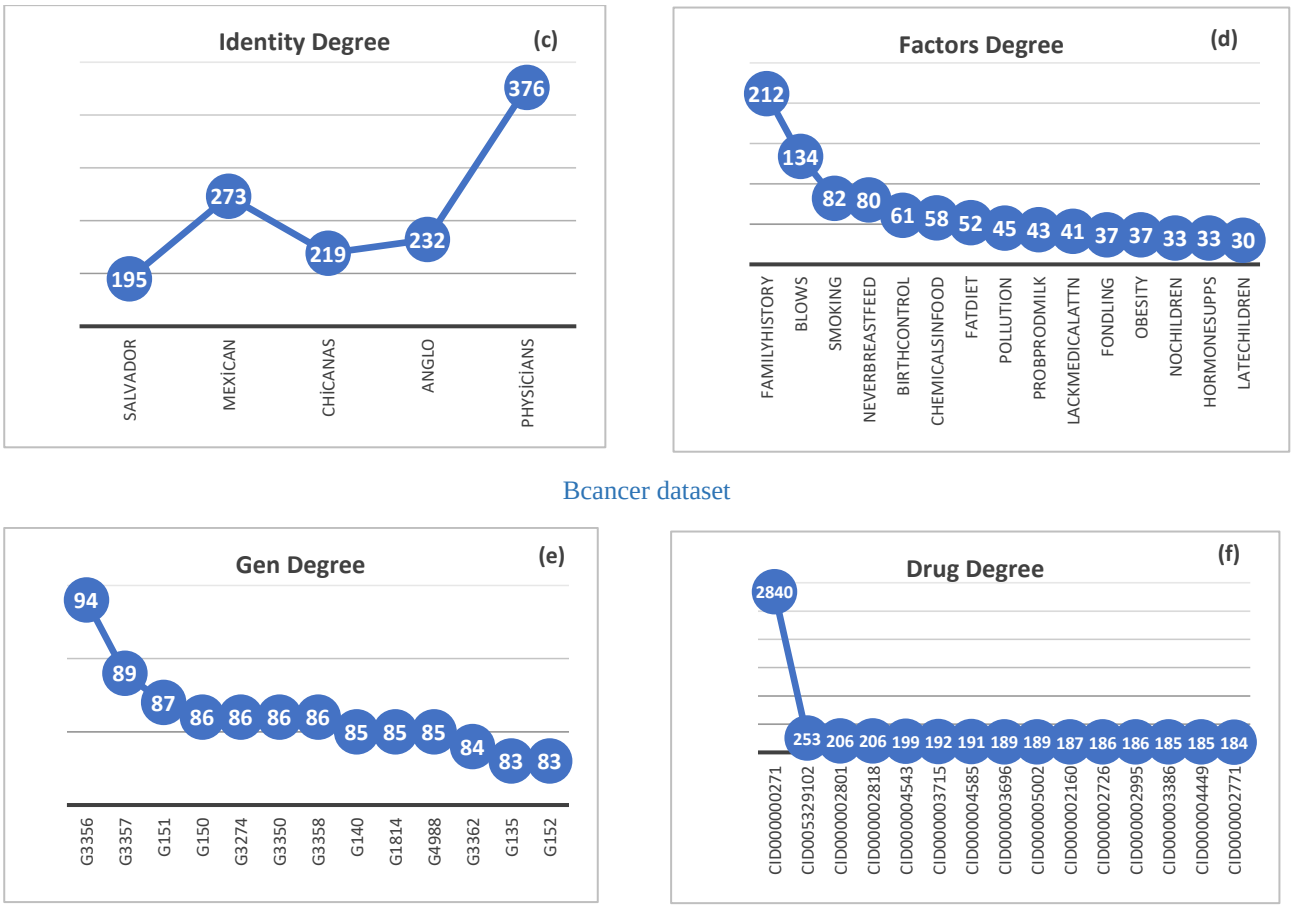


Figure 3. Nodes and hyperedges with the highest degrees in datasets: Almoincidence dataset (a, b), Bcancer dataset (c, d), Drug-Target dataset (e, f)

Degree centrality provides a local measure for centrality. This can be used to compare the results obtained with the proposed method. The top nodes and hyperedges with the highest degree are shown in Figure 3.

Almoincidence

This dataset contains the actors/actress and the movies they acted. The network consists of 94 nodes and 13 hyperedges. In addition, for this dataset, measurements were made with closeness, betweenness, eigenvector and pagerank, which are popular centrality measurements used to give an idea. It is important that which similarity matrix was used when calculating these centralities. There exist challenges to overcome while utilizing the jaccard, for example, a fully connected structure can occur. In this case, centrality measures may not be discriminative. In order to pass this problem, the mean of the jaccard values can be taken as the threshold value and the similarity values can be converted to 0 and 1 accordingly. It can be done with traditional centrality measurements over the generated neighborhood matrix. When tried, traditional centrality measurements with jaccard for this dataset lost data, the most central node such as JoSalcedo could not be found. When considered logically and tried, it was seen that the intersection should be chosen as the most suitable matrix for these traditional centrality measurements. The results obtained according to the intersection are given in the Table 5.

When the results are compared with the ranking obtained by the proposed method, they show very close similarity.

Table 5. Top 15 centrality measure values for actors/actress

betweenness		closeness		eigenvector		pagerank	
JoSalcedo	0.219234	JoSalcedo	1	JoSalcedo	0.2969519	JoSalcedo	0.043025
PeAlmodovar	0.219234	PeAlmodovar	1	PeAlmodovar	0.2969519	PeAlmodovar	0.043025
AgAlmodovar	0.0585377	AgAlmodovar	0.75	AgAlmodovar	0.2275763	AgAlmodovar	0.0286903
EsGarcia	0.0585377	EsGarcia	0.75	EsGarcia	0.2275763	EsGarcia	0.0286903
CaMaura	0.0385087	CaMaura	0.6888889	CaMaura	0.1912972	CaMaura	0.0239346
ALFernandez	0.0261302	ALFernandez	0.6503497	ChLampreave	0.1670298	ALFernandez	0.0207551
MaParedes	0.021283	ChLampreave	0.6413793	KiManver	0.1649636	ChLampreave	0.0194753
KiManver	0.0196501	KiManver	0.6369863	ALFernandez	0.1645402	MaParedes	0.0190321
ChLampreave	0.0193637	MaParedes	0.6369863	MaParedes	0.1637093	KiManver	0.018881
JuSerrano	0.0187646	JuSerrano	0.6283784	JuSerrano	0.1550742	JuSerrano	0.0181879
CeRoth	0.0179644	CeRoth	0.615894	RdPalma	0.1500216	CeRoth	0.0172539
RdPalma	0.0100128	RdPalma	0.6118421	AnBanderas	0.1421031	RdPalma	0.0162173
AnBanderas	0.0084787	AnBanderas	0.6038961	CeRoth	0.1348246	AnBanderas	0.0154106
NaMartinez	0.007604	BeBonezi	0.5923567	BeBonezi	0.1307885	ViAbril	0.0143355
ViAbril	0.0065776	BiAndersen	0.5923567	NaMartinez	0.1268467	BiAndersen	0.0142692

Table 6. The ranking of hyperedge and node centralities for Almoincidence

Columns (Hyperedges) entropy values

degree		e_degree		e_union		e_intersection		e_jaccard	
Film1	16	Film1	4	Film2	3.58378	Film6	3.54583	Film6	3.52849
Film13	15	Film13	3.90689	Film1	3.58314	Film2	3.53608	Film2	3.52568
Film9	15	Film9	3.90689	Film11	3.58251	Film1	3.52729	Film1	3.50982
Film7	15	Film7	3.90689	Film3	3.58247	Film11	3.52331	Film11	3.50375
Film12	14	Film12	3.80735	Film6	3.58228	Film3	3.5106	Film3	3.48715
Film11	14	Film11	3.80735	Film12	3.5819	Film7	3.50841	Film7	3.47162
Film10	14	Film10	3.80735	Film10	3.5817	Film13	3.4771	Film13	3.43673
Film8	14	Film8	3.80735	Film9	3.58146	Film9	3.47038	Film9	3.4358
Film6	14	Film6	3.80735	Film13	3.5814	Film4	3.45854	Film10	3.42041
Film5	14	Film5	3.80735	Film8	3.58103	Film8	3.45414	Film4	3.41549
Film4	14	Film4	3.80735	Film7	3.58092	Film10	3.45363	Film12	3.40769
Film3	13	Film3	3.70044	Film4	3.58062	Film5	3.44618	Film8	3.40231

Rows (Nodes) entropy values

e_degree		e_intersection		e_union		e_jaccard	
JoSalcedo	3.70044	JoSalcedo	6.08382	JoSalcedo	6.53916	JoSalcedo	6.08382
PeAlmodovar	3.70044	PeAlmodovar	6.08382	PeAlmodovar	6.53916	PeAlmodovar	6.08382
AgAlmodovar	3	AgAlmodovar	5.55895	AgAlmodovar	6.52975	AgAlmodovar	5.63864
EsGarcia	3	EsGarcia	5.55895	EsGarcia	6.52975	EsGarcia	5.63864
CaMaura	2.58496	CaMaura	5.37144	CaMaura	6.51482	CaMaura	5.48868
ALFernandez	2.32193	ALFernandez	5.18042	ChLampreave	6.50696	ALFernandez	5.30048
ChLampreave	2.32193	KiManver	5.13573	ALFernandez	6.50025	MaParedes	5.26779
AnBanderas	2	MaParedes	5.1334	RdPalma	6.49291	KiManver	5.26249
JuSerrano	2	ChLampreave	5.10723	AnBanderas	6.49226	ChLampreave	5.23232
KiManver	2	JuSerrano	5.07424	KiManver	6.48835	JuSerrano	5.1861
MaParedes	2	CeRoth	5.03418	MaParedes	6.48745	CeRoth	5.03378
RdPalma	2	RdPalma	4.86	JuSerrano	6.48614	RdPalma	4.99562
AfBeato	1.58496	AnBanderas	4.81727	BeBonezi	6.45845	AnBanderas	4.93421

The similarities of the methods can be explained by comparing the number of order-independent common nodes for the first 15 nodes. As illustrated in Figure 4 that jaccard entropy centrality and betweenness centrality show similar behavior. And, intersection entropy centrality and eigenvector centrality show similar behavior.

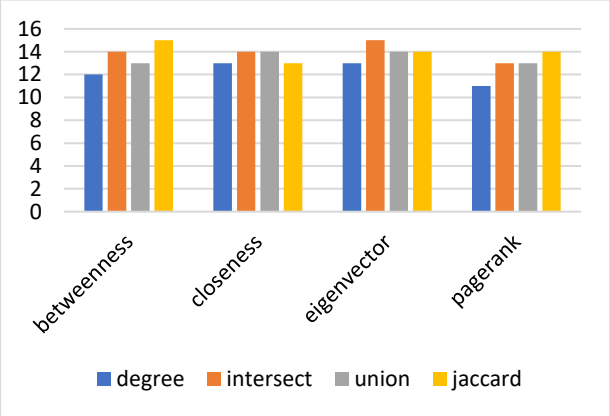


Figure. 4. Comparison of traditional and proposed methods over Almoincidence nodes

When Table 6 is examined, intersection and jaccard gave similar results for hyperedges. There are 13 hyperedges in total. When we examine Film6, which seems to be the most effective in intersection and jaccard, it hosts nodes (PeAlmodovar, ALFernandez, AgAlmodovar, AnBanderas, BiAndersen, CaMaura, EsGarcia, EuPoncela, HeLine, JoSalcedo, MAPCAmpos, MaVelasco, MiMolina, NaMartinez). Those are the elements of 3 or more hyperedges. In other words, they acted in 3 or more movies. That's why the centrality of Film6 turned out to be high. As the number of nodes or hyperedges increases and their similarity decreases, the ranking will be more discriminative.

Bcancer

Breast cancer is the most common cancer among Hispanic/Latino women and is the leading cause of cancer death among this group in the United States [41]. Bcancer data from Ucinet database is in weighted hypergraph structure. The results of the proposed method were evaluated by including weighted values in the calculation. The causes of cancer were analyzed over the created hypergraph.

The network consists of 33 nodes and 5 hyperedges. Accordingly, the values obtained by the proposed method are given in the Table 7. Smoking was the central node in all calculations. Jaccard and intersect gave similar results again.

Table 7. The ranking of hyperedge and node centralities for Bcancer

Columns (Hyperedges) entropy values									
degree		**e_degree**		**e_intersection**		**e_union**		**e_jaccard**	
Physicians	376	Mexican	3.51035	Anglo	1.9699	Physicians	1.99488	Anglo	1.92586
Mexican	273	Salvador	3.48718	Mexican	1.94908	Salvador	1.99338	Mexican	1.91209
Anglo	232	Physicians	3.38983	Chicanas	1.94566	Mexican	1.99242	Physicians	1.90489
Chicanas	219	Chicanas	3.35887	Physicians	1.93291	Chicanas	1.99199	Chicanas	1.90475
Salvador	195	Anglo	3.19052	Salvador	1.83659	Anglo	1.9917	Salvador	1.79108
Rows (Nodes) entropy values									
e_degree		**e_jaccard**		**e_intersection**		**e_union**			
Smoking	2.3049	Smoking	4.76513	Smoking	4.76513	Smoking	5		
Birthcontrol	1.9647	Familyhistory	4.76513	Familyhistory	4.76513	Familyhistory	5		
Neverbreastfeed	1.8728	Birthcontrol	4.63419	Birthcontrol	4.59708	Birthcontrol	4.99131		
Familyhistory	1.8077	Blows	4.47119	Blows	4.44228	Blows	4.99114		
Blows	1.7997	Neverbreastfeed	4.47119	Neverbreastfeed	4.44228	Neverbreastfeed	4.99114		
Fatdiet	1.4059	Nochildren	4.40815	Nochildren	4.40386	Fatdiet	4.96807		
Chemicalsinfood	0.9991	Fatdiet	4.37386	Fatdiet	4.30423	Chemicalsinfood	4.96069		
Fibrocystic	0.995	Fibrocystic	4.11798	Fibrocystic	4.0958	Lackmedicalattn	4.96069		
Pollution	0.9825	Chemicalsinfood	4.00621	Chemicalsinfood	4.00182	Pollution	4.95784		
Illegaldrugs	0.9819	Lackmedicalattn	4.00621	Lackmedicalattn	4.00182	Nochildren	4.93597		
Nochildren	0.9673	Probprodmilk	3.95958	Probprodmilk	4.00163	Probprodmilk	4.93479		
Fondling	0.9569	Fondling	3.95958	Fondling	4.00163	Fondling	4.93479		
Lackmedicalattn	0.9474	Lackhygiene	3.95958	Lackhygiene	4.00163	Lackhygiene	4.93479		

Lackhygiene	0.9457	Illegaldrugs	3.95958	Illegaldrugs	4.00163	Illegaldrugs	4.93479
Probprodmilk	0.9103	Pollution	3.78052	Pollution	3.72193	Fibrocystic	4.92809
Implants	0	Breast-Feeding	3.54759	Breast-Feeding	3.70044	Largebreasts	4.91239
Wildlife	0	Alcohol	3.54759	Alcohol	3.70044	Caffeine	4.91239
Abortions	0	Obesity	3.3674	Obesity	3.58496	Radiation	4.90318
Dirtywork	0	Hormonesupps	3.3674	Hormonesupps	3.58496	Diet	4.90318
Breast-Feeding	0	Latechildren	3.3674	Latechildren	3.58496	Justhappens	4.90318
Alcohol	0	Cancerhistory	3.3674	Cancerhistory	3.58496	Breast-Feeding	4.90215
Largebreasts	0	Age	3.3674	Age	3.58496	Alcohol	4.90215
Caffeine	0	Ethnicity	3.3674	Ethnicity	3.58496	Implants	4.87561
Radiation	0	Earlymenses	3.3674	Earlymenses	3.58496	Wildlife	4.87561
Diet	0	Implants	3.22661	Implants	3.45943	Abortions	4.87561
Justhappens	0	Wildlife	3.22661	Wildlife	3.45943	Dirtywork	4.87561
Obesity	0	Abortions	3.22661	Abortions	3.45943	Obesity	4.84096
Hormonesupps	0	Dirtywork	3.22661	Dirtywork	3.45943	Hormonesupps	4.84096
Latechildren	0	Largebreasts	3.11291	Largebreasts	3.32193	Latechildren	4.84096
Cancerhistory	0	Caffeine	3.11291	Caffeine	3.32193	Cancerhistory	4.84096
Age	0	Radiation	3.04719	Radiation	3.32193	Age	4.84096
Ethnicity	0	Diet	3.04719	Diet	3.32193	Ethnicity	4.84096
Earlymenses	0	Justhappens	3.04719	Justhappens	3.32193	Earlymenses	4.84096

Drug-Target Interaction Network

Drug targets are molecules that play a critical role in drug transport, delivery, or activation. For drug target discovery, drug design, drug docking or screening, drug metabolism prediction, drug interaction prediction, and general pharmaceutical research, drug target information is used [40]. The network consists of 284 nodes and 3648 hyperedges.

Table 8. The ranking of hyperedge and node centralities for Drug-Target Interaction

Columns (Hyperedges) entropy values							
e_degree		**e_jaccard**		**e_intersection**		**e_union**	
G3356	6.5546	G55244	11.6245	G7155	11.5844	G3356	11.8323
G3357	6.4757	G1576	11.5811	G7018	11.583	G3357	11.8323
G151	6.4429	G4215	11.5575	G6326	11.5604	G3358	11.8323
G150	6.4263	G5604	11.5575	G1021	11.5576	G135	11.8322
G3274	6.4263	G51347	11.5519	G140469	11.5576	G140	11.8322
G3350	6.4263	G57551	11.5519	G1612	11.5576	G150	11.8322
G3358	6.4263	G6198	11.5519	G1613	11.5576	G151	11.8322
G140	6.4094	G146802	11.5513	G23604	11.5576	G152	11.8322
G1814	6.4094	G7018	11.5488	G23678	11.5576	G153	11.8322
G4988	6.4094	G120892	11.5443	G27330	11.5576	G155	11.8322
G3362	6.3923	G6790	11.5399	G340156	11.5576	G1814	11.8322
G135	6.375	G316	11.5319	G5585	11.5576	G3274	11.8322
G152	6.375	G1021	11.5286	G5588	11.5576	G3350	11.8322
G153	6.3576	G140469	11.5286	G5592	11.5576	G3351	11.8322
G155	6.3576	G1612	11.5286	G57118	11.5576	G3362	11.8322
Rows (Nodes) entropy values							
e_degree		**e_jaccard**		**e_intersection**		**e_union**	
CID000000271	11.4717	CID000002812	6.9733	CID000002812	7.00897	CID000000271	8.14421
CID005329102	7.98299	CID000004932	6.9457	CID000003198	6.96694	CID005329102	8.04102

CID000002801	7.6865	CID000002156	6.9412	CID000004601	6.88436	CID000002801	8.02022
CID000002818	7.6865	CID000003198	6.8822	CID000004932	6.88323	CID000002818	8.02009
CID000004543	7.63662	CID000003117	6.8818	CID000002156	6.88286	CID000004543	8.01574
CID000003715	7.58496	CID000002756	6.796	CID000003957	6.86086	CID000003715	8.01144
CID000004585	7.57743	CID000002153	6.6629	CID000002267	6.82818	CID000004585	8.00802
CID000003696	7.56224	CID000005035	6.6602	CID000124087	6.81632	CID000005002	8.00627
CID000005002	7.56224	CID000005253	6.5964	CID000005035	6.8	CID000003696	8.00559
CID000002160	7.54689	CID000004601	6.573	CID000002099	6.79511	CID000002726	8.00406
CID000002726	7.53916	CID000124087	6.5128	CID000005530	6.77392	CID000002160	8.00377
CID000002995	7.53916	CID000002267	6.4922	CID000004473	6.77196	CID000002995	8.00259
CID000003386	7.53138	CID000004543	6.4908	CID000003042	6.73112	CID000036811	8.00257
CID000004449	7.53138	CID000002818	6.4876	CID000004178	6.70111	CID000002520	8.00232
CID000002771	7.52356	CID000002801	6.4826	CID000003117	6.69876	CID000004449	8.0018

When gene centrality was examined, degree and union showed similar behavior. The number of genes with similar values in intersection was high. In this dataset, the difference was more for jaccard and intersection for ranking order as seen in Table 8. G7018, G1021, G140469, G1612 are the common genes in the top 15 for jaccard and intersection.

The degree entropy value of the drug CID000000271 was higher than the others. This became the highest influential node or hyperedge for degree and union. The most influential for jaccard and intersection was CID000002812. The number of common elements in the top 15 for the jaccard and intersection results indicates the similarity of the calculation method.

Although CID000000271's rating is too high, it did not appear in the top positions in jaccard and intersection. This shows that when global measurements are desired, intersection and jaccard similarity parameters can be used.

Discussions

Since hypergraphs contain multidimensional information, more factors are effective in the measurement of centrality. It allows the inclusion of relevant data with any number of conditions. Therefore, a multidimensional and detailed centrality measure is obtained. Changing the representation of nodes and hyperedges allows centralities to be viewed from two angles. In addition, measuring weights over similarity is a measure obtained by considering the whole network. It can be a guide in analyzing relationships that cannot be seen, especially as the data grows. The computational cost is also lower in hypergraphs. Hypergraphs are more flexible than graphs, which can make them easier to use in a variety of applications. More is said with less data. Centrality calculation for all data with multidimensional interaction can be easily done in this way. The entropy used for centrality provided a convenient solution for interpreting the local and global potential of the node or hyperedge in the network. Although the most influential nodes are located at the center of the network, it is difficult to distinguish what type of influence they exert on the network [42], [43]. The effect type of nodes on the network was tried to be understood with the results of different entropy calculations. The choice of similarity type used for entropy indicated the direction of the effect.

Conclusions

This paper presented a new efficient centrality measurement for the nodes of a hypergraph. A new method for measuring centrality over entropy using the neighborhood-based similarity values of nodes and hyperedges. In order to investigate the efficiency of the proposed method, we carried out experiments on several networks and demonstrated the effectiveness. Applying the method to different types of datasets yielded accurate results. It enabled us to reach global and local results according to our similarity choice in hypergraphs. Centrality measurement was performed separately for nodes and hyperedges. Since the hypergraphs were converted into graphs over similarity, the weight value, which in a sense expresses the importance of the edge, was given. Therefore, loss of information was reduced. Since these weights are used in entropy measurements, their effects on the whole system were measured. Measuring centrality in hypergraphs that contain multiple aspects of the interaction environment gave us more information. This analysis can be applied to a wide variety of networks. Hypergraphs can be more scalable than graphs for certain problems, such as clustering and recommender systems. This is because hyperedges can represent multiple relationships between vertices, which can help to reduce the number of vertices and edges that need to be considered. They are able to model complex relationships, they can be more scalable, and they are more flexible. As a result, hypergraphs are becoming increasingly popular in a variety of applications.

Ethics Committee Approval

There is no need to obtain permission from the ethics committee for the article prepared.

Conflict of Interest Statement

There is no conflict of interest with any person / institution in the article prepared. References

References

[1] F. Battiston *et al.*, "Networks beyond pairwise interactions: Structure and dynamics," *Phys. Rep.*, vol. 874, pp. 1–92, Aug. 2020.

[2] S. G. Aksoy, C. Joslyn, C. Ortiz Marrero, B. Praggastis, and E. Purvine, "Hypernetwork science via high-order hypergraph walks," *EPJ Data Sci.*, vol. 9, no. 1, p. 16, Dec. 2020.

[3] K. Bouafia and B. Molnár, "Hypergraph Application on Business Process Performance," *Information*, vol. 12, no. 9, p. 370, Sep. 2021.

[4] E. Busseniers, "General Centrality in a hypergraph," *arxiv*, Mar. 2014.

[5] W. Zhou and L. Nakhleh, "Properties of metabolic graphs: biological organization or representation artifacts?," *BMC Bioinformatics*, vol. 12, no. 1, p. 132, Dec. 2011.

[6] İ. Değirmenci, "Entropy measures and the maximum entropy principle," Hacettepe University, 2011.

[7] A. Ullah, B. Wang, J. Sheng, J. Long, N. Khan, and Z. Sun, "Identifying vital nodes from local and global perspectives in complex networks," *Expert Syst. Appl.*, vol. 186, p. 115778, Dec. 2021.

[8] S. Feng *et al.*, "Hypergraph models of biological networks to identify genes critical to pathogenic viral response," *BMC Bioinformatics*, vol. 22, no. 1, p. 287, Dec. 2021.

[9] D. Zhou, J. Huang, and B. Schölkopf, "Learning with hypergraphs: Clustering, classification, and embedding," in *Advances in Neural Information Processing Systems*, 2007.

[10] P. Bonacich, A. Cody Holdren, and M. Johnston, "Hyper-edges and multidimensional centrality," *Soc. Networks*, vol. 26, no. 3, pp. 189–203, Jul. 2004.

[11] S. Bai, F. Zhang, and P. H. S. Torr, "Hypergraph Convolution and Hypergraph Attention," Jan. 2019.

[12] E. Ramadan, A. Tarafdar, and A. Pothen, "A hypergraph model for the yeast protein complex network," in *18th International Parallel and Distributed Processing Symposium, 2004. Proceedings.*, 2004, pp. 189–196.

[13] J. Rafferty *et al.*, "Ranking sets of morbidities using hypergraph centrality," *J. Biomed. Inform.*, vol. 122, 2021.

[14] A. R. Benson, "Three Hypergraph Eigenvector Centralities," *SIAM J. Math. Data Sci.*, vol. 1, no. 2, 2019.

[15] C. Cui, X. Li, C. Zhang, W. Guan, and M. Wang, "Temporal-Relational hypergraph tri-Attention networks for stock trend prediction," *Pattern Recognit.*, vol. 143, p. 109759, Nov. 2023.

[16] J. Wang *et al.*, "Dynamic weighted hypergraph convolutional network for brain functional connectome analysis," *Med. Image Anal.*, vol. 87, p. 102828, Jul. 2023.

[17] S. Klamt, U.-U. Haus, and F. Theis, "Hypergraphs and Cellular Networks," *PLoS Comput. Biol.*, vol. 5, no. 5, p. e1000385, May 2009.

[18] Y. Yang, J. Pei, and A. Al-Barakati, "Measuring in-network node similarity based on neighborhoods: a unified parametric approach," *Knowl. Inf. Syst.*, vol. 53, no. 1, pp. 43–70, Oct. 2017.

[19] S. Li, J. Huang, Z. Zhang, J. Liu, T. Huang, and H. Chen, "Similarity-based future common neighbors model for link prediction in complex networks," *Sci. Rep.*, vol. 8, no. 1, p. 17014, Dec. 2018.

[20] İ. Tuğal, M. Kaya, and T. Tuncer, "Link prediction in disease and drug networks," in *6-th International Conference of Advanced Computer Systems and Networks: Design and Application (ACSN 2013)*, 2013, pp. 46–50.

[21] C. E. Shannon, "A Mathematical Theory of Communication," *Bell Syst. Tech. J.*, 1948.

[22] İ. Tuğal and A. Karcı, "Comparisons of Karcı and Shannon entropies and their effects on centrality of social networks," *Phys. A Stat. Mech. its Appl.*, vol. 523, pp. 352–363, 2019.

[23] Y. Song and Y. Deng, "Entropic Explanation of Power Set," *Int. J. Comput. Commun. Control*, vol. 16, no. 4, Aug. 2021.

[24] C. Hark and A. Karcı, "Karcı summarization: A simple and effective approach for automatic text summarization using Karcı entropy," *Inf. Process. Manag.*, vol. 57, no. 3, p. 102187, May 2020.

[25] Y. Deng, "Deng entropy," *Chaos, Solitons and Fractals*, vol. 91, pp. 549–553, 2016.

[26] V. Chellappan, K. M. Sivalingam, and K. Krithivasan, "A Centrality Entropy Maximization Problem in Shortest Path Routing Networks," *Comput. Networks*, vol. 104, pp. 1–15, Jul. 2016.

[27] Z. Qiu, T. Fan, M. Li, and L. Lü, "Identifying vital nodes by Achlioptas process," *New J. Phys.*, vol. 23, no. 3, p. 033036, Mar. 2021.

[28] C. Hark, T. Uçkan, E. Seyyarer, and A. Karcı, "An Unsupervised Approach Based on Node Centralization for Text Summarization," *Bitlis Eren Univ. J. Sci.*, vol. 8, no. 3, pp. 1109–1118, Sep. 2019.

[29] L. C. Freeman, "Centrality in social networks conceptual clarification," *Soc. Networks*, vol. 1, no. 3, pp. 215–239, 1978.

[30] L. C. Freeman, "A Set of Measures of Centrality Based on Betweenness," *Sociometry*, vol. 40, no. 1, p. 35, Mar. 1977.

[31] P. Bonacich, "Some unique properties of eigenvector centrality," *Soc. Networks*, vol. 29, no. 4, pp. 555–564, 2007.

[32] S. Brin and L. Page, "The anatomy of a large-scale hypertextual Web search engine," *Comput. Networks ISDN Syst.*, vol. 30, no. 1–7, pp. 107–117, Apr. 1998.

[33] J. P. Scott, *Social Network Analysis: A Handbook*. 2000.

[34] E. Estrada, *The Structure of Complex Networks*, vol. 45, no. 2. 2011.

[35] P. Hu, W. Fan, and S. Mei, "Identifying node importance in complex networks," *Phys. A Stat. Mech. its Appl.*, vol. 429, pp. 169–176, 2015.

[36] Z. W. Liang and J. P. Li, "Identifying and ranking influential spreaders in complex networks," in *2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2014*, 2014, pp. 393–396.

[37] S. Wang, Y. Du, and Y. Deng, "A new measure of identifying influential nodes: Efficiency centrality," *Commun. Nonlinear Sci. Numer. Simul.*, vol. 47, pp. 151–163, 2017.

[38] Y. Yu, B. Zhou, L. Chen, T. Gao, and J. Liu, "Identifying Important Nodes in Complex Networks Based on Node Propagation Entropy," *Entropy*, vol. 24, no. 2, p. 275, Feb. 2022.

[39] S. P. Borgatti, M. G. Everett, and L. C. Freeman, "UCINET 6 For Windows: Software for Social Network Analysis, Analytic Technologies, Harvard, MA.," *Anal. Technol.*, 2002.

[40] M. Zitnik, R. Sosi, S. Maheshwari, and J. Leskovec, "BioSNAP Datasets: Stanford Biomedical Network Dataset Collection," 2018. [Online]. Available: https://snap.stanford.edu/biodata/datasets/10015/10015-ChG-TargetDecagon.html.

[41] E. J. Power, M. L. Chin, and M. M. Haq, "Breast Cancer Incidence and Risk Reduction in the Hispanic Population," *Cureus*, Feb. 2018.

[42] Y. Liu, B. Wei, Y. Du, F. Xiao, and Y. Deng, "Identifying influential spreaders by weight degree centrality in complex networks," *Chaos, Solitons & Fractals*, vol. 86, pp. 1–7, May 2016.

[43] T. Bian and Y. Deng, "A new evidential methodology of identifying influential nodes in complex networks," *Chaos, Solitons and Fractals*, vol. 103, pp. 101–110, 2017.