

Analyse des mesures de hiérarchie et de centralité dans les grands graphes de terrain

Stephany Rajeh *, Marinette Savonnet *, Éric Leclercq *, Hocine Cherifi *

*Laboratoire d'Informatique de Bourgogne EA 7534 - Université de Bourgogne
stephany.rajeh@u-bourgogne.fr

Résumé. L'identification des nœuds influents dans les grands graphes de terrain est cruciale dans de nombreuses applications. Les mesures de hiérarchie et de centralité sont les deux approches principales pour quantifier la notion d'influence d'un nœud. Bien que de nombreux travaux concernent les relations entre les différentes mesures de centralité, aucune étude n'a été menée à ce jour pour caractériser les relations entre les mesures de hiérarchie et de centralité. Dans cet article, une évaluation comparative est réalisée en utilisant un ensemble de réseaux du monde réel provenant de divers domaines. Les résultats indiquent que les mesures de hiérarchie et de centralité sont plus ou moins complémentaires selon les propriétés topologiques des réseaux. Plus précisément, la redondance des informations entre les deux types de mesures est fortement liée à la densité et à la transitivité des réseaux. De plus, les mesures de hiérarchie basées sur la notion d'imbrication de sous-réseaux sont les plus orthogonales aux mesures de centralité étudiées.

1 Introduction

Les systèmes tels que les transports, l'Internet, le World Wide Web, les interactions biologiques et chimiques et les connexions sociales ne sont que quelques exemples de systèmes ayant de nombreuses entités inter-connectées existant dans le monde réel. Les réseaux fournissent une représentation robuste de ces systèmes en modélisant les entités par des nœuds et leurs interactions par des liens. L'identification des nœuds influents au sein de ces réseaux est un problème crucial. En effet, ceux-ci permettent de contrôler la dynamique des épidémies, de repérer les utilisateurs influents pour les campagnes de marketing, de délier les réseaux terroristes, et de découvrir des protéines essentielles (Lü et al., 2016). Pour plus de détails sur les principales applications le lecteur pourra consulter Barabási et al. (2016).

L'une des principales façons d'identifier les nœuds influents passe par les mesures de centralité. Celles-ci mesurent la capacité d'un nœud à influencer d'autres nœuds au niveau du voisinage local ou plus globalement au niveau de l'ensemble des nœuds du réseau. On peut distinguer trois types majeurs de mesures de centralité : celles basées sur le voisinage local, celles basées sur les chemins dans le graphe et celles définies par un processus itératif (Lü et al., 2016). Ces mesures peuvent également être combinées et incorporer des informations

sur la structure communautaire des réseaux (Sciarra et al., 2018; Ghalmane et al., 2019a,b; Cherifi et al., 2019). La hiérarchie est un autre moyen d'identifier les nœuds importants au sein d'un réseau. Les structures hiérarchiques peuvent se manifester dans les réseaux écologiques, les réseaux de transport, les réseaux biologiques, les réseaux sociaux et bien d'autres domaines (Zafeiris et Vicsek, 2017). La hiérarchie peut également être considérée sous différents aspects, à savoir la hiérarchie emboîtée et la hiérarchie de flux. La première est basée sur la décomposition hiérarchique du réseau tandis que la seconde est basée sur la hiérarchie des flux de ressources (Lane, 2006).

Étant donné que hiérarchie et centralité traduisent toutes deux la notion d'importance, il est primordial d'examiner les relations qui existent entre ces deux notions. De nombreuses études ont été menées sur les interactions entre les diverses mesures de centralité (Li et al., 2015; Ronqui et Travieso, 2015; Schoch et al., 2017; Oldham et al., 2019). Néanmoins, les relations entre mesures de hiérarchie et de centralité n'ont pas fait l'objet d'études à notre connaissance. Les résultats des travaux que nous avons menés apportent de nombreux enseignements quant aux relations entre ces deux types de mesure d'influence et les propriétés topologiques des réseaux (Rajeh et al., 2020).

Trois questions principales sont abordées :

- Est-ce que les mesures de hiérarchie et de centralité fournissent des informations similaires ?
- Quel est l'impact de la topologie du réseau sur les relations entre ces deux types de mesures ?
- Quelles sont les mesures de hiérarchie et de centralité les plus orthogonales ?

Pour répondre à ces questions, six mesures de centralité et quatre mesures de hiérarchie parmi les plus populaires ont été sélectionnées. Les expérimentations ont porté sur 28 réseaux du monde réel de diverses origines (réseaux sociaux, technologiques, information, etc.). Le détail des propriétés macroscopiques de ces réseaux est donné dans (Rajeh et al., 2020). Les résultats des expérimentations mettent en évidence les interactions entre les mesures de hiérarchie et de centralité et les propriétés topologiques macroscopiques des réseaux telles que la densité et la transitivité. Cela permet de distinguer les situations dans lesquelles les mesures peuvent se substituer pour des raisons d'efficacité, de celles où elles peuvent être combinées afin d'obtenir des mesures d'influence plus performantes.

2 Présentation des mesures et des données

Les six mesures de centralité utilisées sont représentatives des trois catégories les plus courantes. Ainsi le degré et la centralité locale sont des mesures qui caractérisent l'influence d'un nœud dans un voisinage à l'ordre 1 et 2 respectivement. La centralité d'intermédiarité (Betweenness) et la centralité de proximité (Current-flow Closeness) reposent sur les plus courts chemins dans le graphe. Deux mesures de centralité reposant sur un processus itératif sont aussi utilisées (PageRank, Katz). Leur définition est reportée dans le tableau 2. En ce qui concerne les mesures de hiérarchie définies dans le tableau 1, deux d'entre elles sont basées sur la notion d'imbrication de sous-réseaux (k -core, k -truss). LRC (Local Reaching Centrality) est quant à elle basée sur la notion de flux à travers les niveaux hiérarchiques. Finalement, "Triangle Participation" intègre ces deux notions. Pour des informations détaillées, on peut se référer à (Rajeh et al., 2020).

Mesure de hiérarchie	Définition
k -core	$\alpha_c(v_i) = k_{max}^c - (c(v_i) - 1)$
k -truss	$\alpha_t(v_i) = k_{max}^t - k_{min}^t - (t(v_i) - 1)$
LRC	$\alpha_l(v_i) = \frac{1}{(N-1)} \sum_{j=1}^N \frac{1}{d(v_i, v_j)}$
Triangle participation	$\alpha_{tp}(v_i) = k_{max}^{tp} + 1 - Tp(v_i)$

TAB. 1 – Définition des mesures de hiérarchie (α_i). $c(v_i)$ = numéro de core du nœud v_i . k_{max}^c = degré maximal induisant le sous-graphe maximal de k -core du graphe G . $t(v_i)$ = numéro de truss du nœud v_i . k_{max}^t = nombre maximum de triangles induisant un sous-graphe maximal de k -truss du graphe G . k_{min}^t = nombre minimum de triangles de telle sorte que $k_{min}^t \geq 2$. N = nombre total de nœuds. $d(v_i, v_j)$ = distance entre les nœuds v_i et v_j . k_{max}^{tp} = nombre total de triangles dans un graphe G . $Tp(v_i)$ = nombre de triangles auquel nœud v_i participe.

Les graphes de terrain utilisés proviennent de domaines variés. On compte ainsi des réseaux sociaux impliquant des animaux, des humains et des comptes/pages en ligne. Un autre domaine est la biologie où les liens représentent des réactions physiques ou chimiques entre des entités biologiques. Les réseaux de collaboration sont également utilisés. Les réseaux d'infrastructures comprennent soit des routes en tant que nœuds, soit des frontières, soit des unités électriques, soit des aéroports. Enfin, divers réseaux impliquent différents types de nœuds tels que les acteurs, les pays, les mots et les systèmes autonomes, respectivement. Le code et les données sont accessible via GitHub¹.

3 Corrélation et similarité des mesures de hiérarchie et de centralité

Pour chacun des réseaux S_i , $i = [1, 28]$, les six mesures de centralité et les quatre mesures de hiérarchie sont calculées sur chaque nœud. Ensuite, la corrélation entre les mesures de hiérarchie et de centralité est calculée. Ceci est fait pour toutes les combinaisons possibles en utilisant 5 mesures d'évaluation (3 mesures de corrélation et 2 mesures de similarité). Les mesures de corrélation retenues sont Pearson (ρ_p), Spearman (ρ_s), et Kendall's Tau (τ_b). La similarité est évaluée en utilisant l'indice de Jaccard (J) et le "Rank-Biased Overlap" (RBO) (Webber et al., 2010). Hormis pour les résultats reportés dans la table 3, chaque mesure est utilisée séparément. La figure 1 illustre des résultats représentatifs pour la corrélation de Spearman des 28 réseaux étudiés. On peut ainsi distinguer 6 catégories. La première carte de corrélation (ex., CS Ph.D.) dénote une faible corrélation entre les mesures de hiérarchie et de centralité. À l'autre extrémité (ex., World Metal Trade), la corrélation est forte entre les deux types de mesures. Entre ces deux extrêmes, il y a une catégorie de réseaux (ex., E. coli) où seuls k -core

1. <https://github.com/StephanyRajeh/HierarchyAndCentralityAnalysis>

Mesure de centralité	Définition
Degree	$\beta_d(v_i) = \frac{1}{(N-1)} \sum_{j=1}^N a_{ij}$
Local	$\beta_l(v_i) = \frac{1}{(N-1)} \sum_{j=1}^N a'_{ij}$
Betweenness	$\beta_b(v_i) = \frac{2}{(N-1)(N-2)} \left(\sum_{s,t \neq i} \frac{\sigma_{v_i}(v_s, v_t)}{\sigma(v_s, v_t)} \right)$
Current-flow closeness	$\beta_c(v_i) = \frac{N}{\sum_{j=1}^N r_{ii} + r_{jj} - 2r_{ij}}$
Katz	$\beta_k(v_i) = \sum_{p=1}^{\infty} \sum_{j=1}^N s^p a_{ij}^p$
PageRank	$\beta_p(v_i) = \frac{1-d}{N} + d \sum_{v_j \in M_i} \frac{\beta_p(v_j)}{k_j}$

TAB. 2 – Définition des mesures de centralité (β_j). a_{ij} = obtenu à partir du voisinage de premier ordre de la matrice d'adjacence A . N = nombre total de nœuds. a'_{ij} = obtenu à partir de la matrice de voisinage de second ordre de la matrice d'adjacence A . $\sigma(v_s, v_t)$ = nombre de chemins les plus courts entre les nœuds v_s et v_t . $\sigma_{v_i}(v_s, v_t)$ = nombre de chemins les plus courts entre les nœuds v_s et v_t qui passent par le nœud v_i . r_{ij} = quantité d'information qui peut être transmise à partir du nœud v_i à v_j à travers toutes les voies possibles. a_{ij}^p = la connectivité du nœud v_i par rapport à tous les autres nœuds obtenue à partir de tous les ordres possibles de la matrice d'adjacence A . p = ordre de la matrice A . s^p = paramètre de pénalité où $s \in [0,1]$. $\beta_p(v_i)$ = Centralité PageRank du nœud v_i . $\beta_p(v_j)$ = Centralité PageRank du nœud v_j . M_i = ensemble des nœuds liés au nœud v_i . k_j = nombre de liens du nœud v_j au nœud v_i . d = paramètre d'amortissement où $d \in [0,1]$.

et LRC présentent une corrélation significative avec les mesures de centralité. La catégorie suivante est caractérisée par une faible corrélation entre le k -core et le k -truss avec les mesures de centralité (ex., Insects) et une corrélation très marquée pour les autres mesures. La catégorie suivante (ex., Physicians) est caractérisée par une faible corrélation entre le k -truss et les mesures de centralité. La cinquième catégorie (ex., Birds) est caractérisée par une forte corrélation entre les mesures de hiérarchie et de centralité à l'exception de la Betweenness. Les résultats obtenus avec les mesures de corrélation de Pearson et Kendall's Tau sont similaires aux résultats précédents.

En ce qui concerne les mesures de similarité, on retrouve des catégories comparables en utilisant l'index de Jaccard. Par contre, le nombre de catégories passe de 6 à 3 avec RBO. Ceci est dû au fait que cette mesure accorde plus de poids aux nœuds de plus grand rang. Ces premiers résultats illustrent donc la diversité des situations.

4 Impact de la topologie des réseaux

Pour étudier l'influence de la topologie des réseaux sur les relations entre les mesures de centralité et de hiérarchie, on ordonne ceux-ci. On observe ensuite l'évolution des propriétés

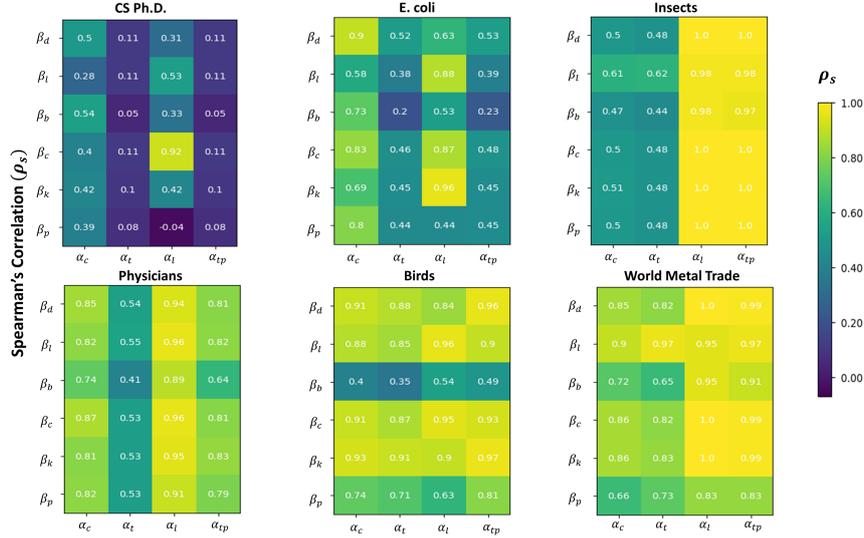


FIG. 1 – Cartes thermiques de la corrélation de Spearman pour les diverses combinaisons de mesures de hiérarchie α_i et de centralité β_j de 6 réseaux représentatifs. Les mesures de hiérarchie sont $\alpha_c = k$ -core, $\alpha_t = k$ -truss, $\alpha_l = LRC$, and $\alpha_{tp} =$ triangle participation. Les mesures de centralité sont $\beta_d =$ Degree, $\beta_l =$ Local, $\beta_b =$ Betweenness, $\beta_c =$ Current-flow Closeness, $\beta_k =$ Katz, and $\beta_p =$ PageRank.

topologiques de base afin de mettre en évidence d'éventuelles relations. Pour ordonner les réseaux, on compte le nombre λ des valeurs des 3 mesures de corrélation qui sont significatives ($\rho_s \geq 0.7 + \rho_p \geq 0.7 + \tau_b \geq 0.7$). Les réseaux sont ordonnés selon le nombre de relations significatives en ordre décroissant. Le tableau 3 résume les résultats obtenus. Notez que les caractéristiques topologiques de base (densité, transitivité et assortativité) sont aussi reportées. Rappelons que la transitivité mesure la proportion de triangles dans un graphe et que l'assortativité mesure la propension des nœuds à posséder des liens avec des nœuds de même degré. L'assortativité est positive lorsque les nœuds de même degré ont tendance à être reliés ensemble. Elle est négative lorsque les nœuds de faible degré ont tendance à être reliés aux nœuds de fort degré. A l'examen du tableau 3, on voit apparaître trois groupes de réseaux. Le premier groupe, qui présente une proportion élevée de corrélation et de similarité significative, se distingue par une densité élevée, une transitivité élevée et une assortativité négative. Le second groupe, qui présente une proportion moyenne de corrélation et de similarité significatives se distingue par une faible densité, une forte transitivité et une assortativité positive. Enfin, le dernier groupe, qui présente une faible proportion de corrélation et de similarité significatives se caractérise par une faible densité, une faible transitivité et une assortativité négative.

Pour évaluer la cohérence des trois groupes, chaque réseau est représenté par son vecteur de mesure d'évaluation. L'algorithme du k -means est alors utilisé pour classifier les réseaux (avec $k=3$). Les résultats de cette expérience sont conformes aux résultats précédents. Au final,

Analyse des mesures de hiérarchie et de centralité

Réseau	N	$ E $	λ	ν	ζ	$k_{nn}(k)$
Adjective Noun	112	425	59	0.068	0.156	-0.129
Zachary Karate Club	34	78	55	0.139	0.255	-0.475
Les Misérables	77	254	55	0.086	0.498	-0.165
World Metal Trade	80	875	54	0.276	0.459	-0.391
U.S. Airports	500	2,980	52	0.023	0.351	-0.267
Madrid Train Bombings	64	243	51	0.120	0.561	0.029
Birds	117	304	51	0.577	0.472	0.062
Mammals	28	235	49	0.716	0.727	-0.004
Physicians	117	465	41	0.068	0.174	-0.084
Facebook Pol. Pages	5,908	41,729	39	0.002	0.301	0.018
Facebook Ego	4,039	88,234	38	0.010	0.519	0.063
Insects	113	4,550	36	0.798	0.785	-0.030
U.S. States	49	107	34	0.090	0.406	0.233
AstroPh	17,903	196,972	34	0.001	0.317	0.201
GrQc	4,158	13,422	33	0.001	0.628	0.639
Adolescent Health	2,539	12,969	29	0.002	0.141	0.231
Reptiles	496	984	26	0.008	0.419	0.342
PGP	10,680	24,316	25	0.0004	0.378	0.238
Retweets Copenhagen	761	1,029	23	0.003	0.060	-0.099
Internet A. Systems	6,474	12,572	20	0.0006	0.009	-0.181
NetSci	379	914	19	0.012	0.430	-0.081
Human Protein	2,217	6,418	19	0.002	0.007	-0.331
E. coli Transcription	329	456	19	0.008	0.023	-0.263
Mouse Vis. Cortex	193	214	13	0.011	0.004	-0.844
Yeast Protein	1,458	1,993	12	0.001	0.051	-0.207
U.S. Power Grids	4,941	6,594	5	0.0005	0.103	0.003
EuroRoads	1,039	1,305	4	0.002	0.035	0.090
CS Ph.D.	1,025	1,043	3	0.001	0.002	-0.253

TAB. 3 – Réseaux du monde réel ordonnés par ordre décroissant selon le nombre de mesures de corrélation significatives (λ). N est le nombre total de nœuds. $|E|$ est le nombre total des liens. Les propriétés topologiques de base des réseaux sont ν , la densité, ζ , la transitivity et $k_{nn}(k)$, l'assortativité.

il apparaît donc que densité et transitivity sont les propriétés topologiques les plus importantes pour prédire les relations entre les mesures de centralité et de hiérarchie dans les réseaux. Pour les réseaux caractérisés par une forte densité et une grande transitivity on peut réduire la complexité des traitements en substituant une mesure de centralité à une mesure de hiérarchie. Pour les réseaux caractérisés par une faible densité et/ou une faible transitivity, on peut combiner les mesures pour améliorer l'identification des nœuds influents.

5 Redondance entre mesures de centralité et de hiérarchie

Afin de vérifier quelles sont les mesures de hiérarchie et de centralité les plus similaires et les plus orthogonales, une expérimentation est menée en utilisant l'ensemble des réseaux. Pour ce faire, la méthode de vote Schulze (Schulze, 2011) est utilisée. Le réseau est considéré comme un électeur, et les valeurs des mesures d'évaluation entre les mesures de hiérarchie et de centralité sont les candidats. La méthode de Schulze produit ainsi une liste classée des combinaisons des mesures de hiérarchie et de centralité. Les classements basés sur les mesures de corrélation sont reportés dans le tableau 4. Les rangs obtenus à partir des autres mesures d'évaluation sont comparables. Dans tous les cas, il apparaît que la combinaison la plus orthogonale est formée par le k -truss et la centralité d'intermédiarité (α_t, β_b) .

Les combinaisons plus orthogonales :	$(\alpha_t, \beta_b), (\alpha_c, \beta_b), (\alpha_l, \beta_p), (\alpha_l, \beta_b), (\alpha_t, \beta_p), (\alpha_{tp}, \beta_c),$ $(\alpha_t, \beta_c), (\alpha_t, \beta_l), (\alpha_c, \beta_p), (\alpha_t, \beta_k), (\alpha_l, \beta_d), (\alpha_t, \beta_d)$
Les combinaisons moins orthogonales :	$(\alpha_{tp}, \beta_p), (\alpha_{tp}, \beta_l), (\alpha_c, \beta_l), (\alpha_{tp}, \beta_b), (\alpha_c, \beta_k), (\alpha_{tp}, \beta_k),$ $(\alpha_l, \beta_k), (\alpha_c, \beta_c), (\alpha_c, \beta_d), (\alpha_{tp}, \beta_d), (\alpha_l, \beta_l), (\alpha_l, \beta_c)$

TAB. 4 – Classement des combinaisons par ordre décroissant entre hiérarchie et centralité (α_i, β_j) basé sur la méthode de Schulze. Les mesures de hiérarchie sont $\alpha_c = k$ -core, $\alpha_t = k$ -truss, $\alpha_l = LRC$, et $\alpha_{tp} =$ triangle participation. Les mesures de centralité sont $\beta_d =$ Degree, $\beta_l =$ Local, $\beta_b =$ Betweenness, $\beta_c =$ Current-flow Closeness, $\beta_k =$ Katz, et $\beta_p =$ PageRank.

6 Conclusion

Afin d'examiner les relations entre les mesures de hiérarchie et de centralité les plus communément utilisées nous avons mené une série d'expérimentations utilisant plusieurs mesures d'évaluation. Il apparaît que la topologie du réseau affecte la relation entre les mesures de hiérarchie et de centralité. Plus précisément, les mesures sont redondantes lorsque le réseau est caractérisé par une forte densité et une forte transitivity. Au contraire, cette redondance décroît lorsque le réseau est caractérisé par une faible densité et/ou une faible transitivity. L'assortativité est moins influente dans les variations observées. Enfin, les mesures de hiérarchie k -core et k -truss sont les plus orthogonales aux mesures de centralité.

Ces résultats impliquent que l'on peut substituer les mesures de centralité aux mesures de hiérarchie lorsqu'un réseau est caractérisé par une forte densité et une forte transitivity à des fins d'efficacité. D'autre part, lorsque les mesures de hiérarchie et de centralité sont complémentaires, leur combinaison permet d'envisager des schémas multidimensionnels pour l'identification des nœuds influents du réseau.

Références

- Barabási, A.-L. et al. (2016). *Network science*. Cambridge university press.
- Cherifi, H., G. Palla, B. K. Szymanski, et X. Lu (2019). On community structure in complex networks : challenges and opportunities. *Applied Network Science* 4(1), 1–35.

- Ghalmane, Z., C. Cherifi, H. Cherifi, et M. El Hassouni (2019a). Centrality in complex networks with overlapping community structure. *Scientific reports* 9(1), 1–29.
- Ghalmane, Z., M. El Hassouni, C. Cherifi, et H. Cherifi (2019b). Centrality in modular networks. *EPJ Data Science* 8(1), 15.
- Lane, D. (2006). Hierarchy, complexity, society. In *Hierarchy in natural and social sciences*, pp. 81–119. Springer.
- Li, C., Q. Li, P. Van Mieghem, H. E. Stanley, et H. Wang (2015). Correlation between centrality metrics and their application to the opinion model. *EPJ B* 88(3), 1–13.
- Lü, L., D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, et T. Zhou (2016). Vital nodes identification in complex networks. *Physics Reports* 650, 1–63.
- Oldham, S., B. Fulcher, L. Parkes, A. Arnatkeviciūtė, C. Suo, et A. Fornito (2019). Consistency and differences between centrality measures across distinct classes of networks. *PLoS one* 14(7), 1–23.
- Rajeh, S., M. Savonnet, E. Leclercq, et H. Cherifi (2020). Interplay between hierarchy and centrality in complex networks. *IEEE Access* 8, 129717–129742.
- Ronqui, J. R. F. et G. Travieso (2015). Analyzing complex networks through correlations in centrality measurements. *J. Stat. Mech. Theory Exp* 2015(5), P05030.
- Schoch, D., T. W. Valente, et U. Brandes (2017). Correlations among centrality indices and a class of uniquely ranked graphs. *Social Networks* 50, 46–54.
- Schulze, M. (2011). A new monotonic, clone-independent, reversal symmetric, and condorcet-consistent single-winner election method. *Social Choice and Welfare* 36(2), 267–303.
- Sciara, C., G. Chiarotti, F. Laio, et L. Ridolfi (2018). A change of perspective in network centrality. *Scientific reports* 8(1), 1–9.
- Webber, W., A. Moffat, et J. Zobel (2010). A similarity measure for indefinite rankings. *ACM Transactions on Information Systems (TOIS)* 28(4), 1–38.
- Zafeiris, A. et T. Vicsek (2017). *Why We Live in Hierarchies? : A Quantitative Treatise*. Springer.

Summary

Identification of influential nodes in complex networks is crucial in many applications. Hierarchy and centrality are two principal approaches to quantify the influence of a node. Although there has been a lot of work about the relationship between centrality measures, to our knowledge there is no study to characterize the interplay between hierarchy and centrality measures. In this paper, based on a collection of real-world networks originating from diverse domains, a comparative investigation is performed. Results indicate that hierarchy and centrality measures are more or less complementary according to the network structure. More precisely, density and transitivity drive the redundancy of information between both types of measures. Additionally, nested hierarchy measures are the most orthogonal to the centrality measures under study.