Supplementary Information

María Pereda[a,b,c,*], Ernesto Estrada[d]

[a] *RWTH Aachen University, Chair for Computational Social Sciences and Humanities, Germany*
[b] *Grupo Interdisciplinar de Sistemas Complejos, Departamento de Matemáticas, Universidad Carlos III de Madrid.*
[c] *Unidad Mixta Interdisciplinar de Comportamiento y Complejidad Social (UMICCS), Spain.*
[d] *Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow, UK.*

1. Example of the use of NMDS for reducing the dimensionality of the communicability angle matrix

In this example at table 1, we report the communicability angle matrix θ for a tree with degree sequence 3,3,2,2,1,1,1,1. Then, using the procedures described in the main text of this work we obtained the communicability angles matrices $\hat{\theta}$ for the four mapping criteria described in the main text. Below each matrix $\hat{\theta}$ we report the RMSE and SE values for the comparison of the corresponding matrix and θ. According to RMSE value the best fit is obtained for the strain method $RMSE = 96.53$. However, a closer look at the matrix $\hat{\theta}$ obtained from this method reveals some worrying signs. In particular, there are some angles very close to zero although in the original communicability angle matrix no angle is below 50°. In contrast, the method based on the eigenvalues of the communicability angle matrices reports that the best fit is obtained for the *metricstress*, method for which such "pathologies" are not observed.

*Corresponding author
Email addresses: maria.pereda@humtec.rwth-aachen.de (María Pereda), ernesto.estrada@strath.ac.uk (Ernesto Estrada)

$$\begin{pmatrix} 0 & 67.0 & 88.2 & 88.2 & 83.4 & 83.4 & 44.1 & 70.9 \\ & 0 & 88.2 & 88.2 & 83.4 & 83.4 & 44.1 & 70.9 \\ & & 0 & 88.2 & 43.3 & 83.4 & 84.2 & 71.9 \\ & & & 0 & 83.6 & 43.3 & 84.2 & 71.9 \\ & & & & 0 & 71.1 & 72.4 & 47.0 \\ & & & & & 0 & 72.4 & 47.0 \\ & & & & & & 0 & 48.3 \\ & & & & & & & 0 \end{pmatrix}$$

$$\begin{pmatrix} 0 & 71.8 & 142.8 & 115.7 & 120.6 & 103.6 & 35.2 & 81.0 \\ & 0 & 95.3 & 95.6 & 126.6 & 130.5 & 47.9 & 150.7 \\ & & 0 & 99.8 & 41.2 & 110.8 & 111.1 & 100.1 \\ & & & 0 & 116.8 & 40.4 & 132.6 & 106.0 \\ & & & & 0 & 98.8 & 110.0 & 59.8 \\ & & & & & 0 & 138.0 & 65.9 \\ & & & & & & 0 & 65.9 \\ & & & & & & & 0 \end{pmatrix}$$

metricstress: $RMSE = 102.49$, $SE = 86.027$

$$\begin{pmatrix} 0 & 66.3 & 99.5 & 99.5 & 124.1 & 124.1 & 52.8 & 150.7 \\ & 0 & 126.4 & 126.4 & 118.6 & 118.6 & 13.5 & 84.4 \\ & & 0 & 106.3 & 28.8 & 111.9 & 123.9 & 98.0 \\ & & & 0 & 111.9 & 28.8 & 124.0 & 98.0 \\ & & & & 0 & 102.6 & 124.0 & 69.3 \\ & & & & & 0 & 124.0 & 69.3 \\ & & & & & & 0 & 97.9 \\ & & & & & & & 0 \end{pmatrix}$$

metricsstress:$RMSE = 99.07$, $SE = 86.387$

$$\begin{pmatrix} 0 & 68.9 & 98.6 & 91.5 & 138.0 & 122.3 & 54.0 & 143.6 \\ & 0 & 104.8 & 153.3 & 98.4 & 132.6 & 43.2 & 84.1 \\ & & 0 & 95.7 & 44.3 & 116.8 & 140.9 & 112.1 \\ & & & 0 & 108.2 & 44.0 & 110.7 & 104.0 \\ & & & & 0 & 96.3 & 138.7 & 67.9 \\ & & & & & 0 & 102.1 & 60.6 \\ & & & & & & 0 & 89.6 \\ & & & & & & & 0 \end{pmatrix}$$

sammon:$RMSE = 103.89$, $SE = 86.553$

$$\begin{pmatrix} 0 & 0.5 & 112.6 & 112.6 & 129.2 & 129.2 & 32.8 & 117.5 \\ & 0 & 112.9 & 112.9 & 129.2 & 129.2 & 32.3 & 117.0 \\ & & 0 & 101.1 & 37.8 & 113.4 & 126.7 & 105.7 \\ & & & 0 & 113.4 & 37.8 & 126.7 & 105.7 \\ & & & & 0 & 100.6 & 118.8 & 68.1 \\ & & & & & 0 & 118.8 & 68.0 \\ & & & & & & 0 & 84.7 \\ & & & & & & & 0 \end{pmatrix}$$

strain:$RMSE = 96.53$, $SE = 87.949$

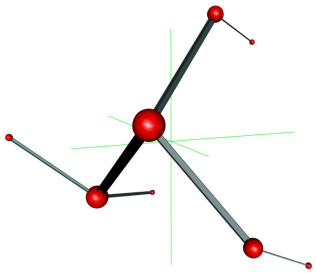

Table 1: Communicability angle matrix (top line) of a tree with degree sequence 3,3,2,2,1,1,1,1. The communicability angle matrices obtained by using NMDS and the following scaling criteria: metricstress (second line, left); metricsstress (second line, right); sammon (third line, left); strain (third line, right). The values of the RMSE and SE for each scaling criterion is given below the angle matrices. The bottom line displays the embedding ito 3D space using the best scaling according to SE, i.e., metricstress.

2. Nonmetric Multidimensional Scaling

The NMDS transforms θ into a matrix of squared Euclidean distances with elements $d_{pq}^2(\theta) = \sum_{r=1}^{n} (\theta_{ps} - \theta_{qs})^2$ which can be obtained by

$$D^{(2)} = \vec{1}\vec{\kappa}^T + \vec{\kappa}\vec{1}^T - 2\theta\theta^T, \tag{1}$$

where $\vec{\kappa} = diag(\theta)$.

The NMDS method works as follows. Starting from a completely random configuration of the n nodes of a network in a 3D space, the distances between the nodes in these random configurations are calculated and regressed using a least-squared method to those in $D^{(2)}$. Then, we calculate the goodness of fit for each configuration by the so-called *stress function* introduced by Kruskal, which is defined by

$$S\left(\theta, \hat{d}\right) = \sqrt{\frac{\sum_{p,q}\left(d_{pq} - \hat{d}_{pq}\right)^2}{\sum_{p,q} d_{pq}^2}}, \tag{2}$$

where $\hat{d}_{pq}$ is the distance between the corresponding nodes in the configuration model. With the goal of minimizing such stress the positions of the nodes in the ordination space are changed in the direction of the steepest descent, which is the direction in which stress changes most rapidily. The procedure is repeated until a minimum (which may be local) is reached or the total number of iterations–here set to 1,000–is reached. Once we have obtained the "optimal" distance matrix $\hat{D}^{(2)}$ we construct the corresponding estimated communicability matrix as follow. We first construct the matrix $J = I - \vec{1}\vec{1}^T/\vec{1}^T\vec{1}$ as the centering matrix and obtain

$$\theta\theta^T = -\frac{1}{2} J \hat{D}^{(2)} J. \tag{3}$$

Then, by using the spectral decomposition of $-\frac{1}{2} J \hat{D}^{(2)} J = Q \Upsilon Q^T$ we finally

obtain

$$\theta = Q\Upsilon^{1/2}. \tag{4}$$

3. k-Means algorithm

K-Means operates by considering the matrix $\hat{\theta}$ as the initial dataset consisting of column vectors which represents the nodes of the graph. That is, the ith column of $\hat{\theta}$ represents an n-dimensional vector representing node i. Then, K-Means generates K non-empty disjoint clusters $C = \{C_1, C_2, \ldots, C_K\}$ around the centroids $c = \{c_1, c_2, \ldots, c_K\}$, by iteratively minimizing the sum [29, 52]

$$W_K = \sum_{k=1}^{K} \sum_{i \in C_K} \sum_{\gamma=1}^{n} \left(\theta_{i\gamma} - c_{k\gamma}\right)^2. \tag{5}$$

4. Cluster Validity Indexes

In the next paragraphs we define the CVIs that we will use in this work. Let c be the center of the network and let c_i be the center of the cluster C_i which has n_i nodes. Let x denotes any node in the network. Then, if the distance between two points p and q is given by $d(p,q)$, the *Calinski-Harabasz index* is defined by [8]

$$CH = \frac{\frac{1}{n_c}\sum_i n_i d^2(c_i, c)}{\frac{1}{n - n_c}\sum_i \sum_{x \in C_i} d^2(x, c_i)}. \tag{6}$$

Let $x \in C_i$ and let $a(x)$ represents the average dissimilarity of the node x to all other nodes in the cluster C_i. If $b(x)$ is the lowest average dissimilarity of x to any other cluster $C_j \neq C_i$, then the *Silhouette index* is defined by [49]

$$S = \frac{1}{n_c}\sum_i \left\{ \frac{1}{n_i} \sum_{x \in C_i} \frac{b(x) - a(x)}{\max\left[a(x), b(x)\right]} \right\}. \tag{7}$$

The *Davies-Bouldin index* is defined as [9]

$$DB = \frac{1}{n_c} \sum_i \max_{j, j \neq i} \left\{ \frac{\frac{1}{n_i} \sum_{x \in C_i} d\left(x, c_i\right) + \frac{1}{n_j} \sum_{x \in C_j} d\left(x, c_j\right)}{d\left(c_i, c_j\right)} \right\}. \tag{8}$$

5. Best" Approach Selection"

5.1. Normalized Mutual Information (NMI)

Here we use a normalization introduced by Strehl and Ghosh for which NMI index is defined as [53]

$$NMI\left(x, y\right) = \frac{I\left(x, y\right)}{\sqrt{H\left(x\right) H\left(y\right)}}, \tag{9}$$

where $I\left(x, y\right)$ is the mutual information between x and y, and $H\left(x\right)$ is the entropy of x. In our case x is a vector which contains a label for the different communities in the ground-truth of the corresponding network and y is a similar vector containing the labels of the clusters obtained by K-means.

5.2. Modularity

The modularity is defined by [44]

$$Q = \frac{1}{4m} \sum_{i,j \in C_k} \left[A_{ij} - \frac{k_i k_j}{2m} \right], \tag{10}$$

where A_{ij} is the corresponding entry of the adjacency matrix, k_i is the degree of the node i, m is the number of edges and the sum is carried out over all communities C_k existing in the network.

6. Tables

N	1	2	3	4	5	6	7	8	9	10
gene	NDRG1	FGF14	NEFH	PPP2R2B	SLC25A22	GABRA1	JPH3	GJB1	UCHL1	DNM2
ref	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]

N	11	12	13	14	15	16	17	18	19
gene	TDP1	SOD1	PARK7	LRRK2	KIF1B	HSPD1	NR4A2	Rab7	SNCAIP
ref.	[11]	[12]	[13]	[14]	[15]	[16]	[17]	[18]	[19]

Table 2: List of references for the findings of cancer activity of genes reported in cluster 1 according to the Table 4 of the main paper.

N	1	2	3	4	5	6	7	8	9	10
gene	ABCA1	ESR1	ALOX5	IL10	IL13	CIITA	PTPRC	BDNF	PLA2G7	CD36
ref	[20]	[21]	[22]	[23]	[24]	[25]	[26]	[27]	[28]	[29]

Table 3: List of references for the findings of cancer activity of genes reported in cluster 51 according to the Table 5 of the main paper.

References

[1] T. P. Ellen, Q. Ke, P. Zhang, M. Costa, Ndrg1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states, Carcinogenesis 29 (1) (2008) 2–8.

[2] T. P. Ellen, Q. Ke, P. Zhang, M. Costa, Ndrg1, a growth and cancer related gene: regulation of gene expression and function in normal and disease states, Carcinogenesis 29 (1) (2008) 2–8.

[3] M. F. Calmon, J. Jeschke, W. Zhang, M. Dhir, C. Siebenks, A. Herrera, H.-C. Tsai, H. M. O'Hagan, E. P. Pappou, C. M. Hooker, T. Fu, K. E. Schuebel, E. Gabrielson, P. Rahal, J. G. Herman, S. B. Baylin, N. Ahuja, Epigenetic silencing of neurofilament genes promotes an aggressive phenotype in breast cancer, Epigenetics 10 (7) (2015) 622–632.

[4] A. Vazquez, D. Kulkarni, L. F. Grochola, G. L. Bond, N. Barnard, D. Toppmeyer, A. J. Levine, K. M. Hirshfield, A genetic variant in a pp2a regulatory

subunit encoded by the ppp2r2b gene associates with altered breast cancer risk and recurrence, International Journal of Cancer 128 (10) (2011) 2335–2343.

[5] C. C. Wong, Y. Qian, X. Li, J. Xu, W. Kang, J. H. Tong, K.-F. To, Y. Jin, W. Li, H. Chen, M. Y. Y. Go, J.-L. Wu, K. W. Cheng, S. S. M. Ng, J. J. Y. Sung, Z. Cai, J. Yu, Slc25a22 promotes proliferation and survival of colorectal cancer cells with kras mutations and xenograft tumor progression in mice via intracellular synthesis of aspartate, Gastroenterology 151 (5) (2016) 945–960.e6.

[6] S. Lee, T. Oh, H. Chung, S. Rha, C. Kim, Y. Moon, Slc25a22 promotes proliferation and survival of colorectal cancer cells with kras mutations and xenograft tumor progression in mice via intracellular synthesis of aspartate, International Journal of Oncology 40.

[7] X. Hu, Y. Kuang, L. Li, H. Tang, Q. Shi, X. Shu, Y. Zhang, F. K. Chan, Q. Tao, C. He, Epigenomic and functional characterization of junctophilin 3 (jph3) as a novel tumor suppressor being frequently inactivated by promoter cpg methylation in digestive cancers, Theranostics 7(7).

[8] S. Sirnes, H. Honne, D. Ahmed, S. A. Danielsen, T. O. Rognum, G. I. Meling, E. Leithe, E. Rivedal, R. A. Lothe, G. E. Lind, Dna methylation analyses of the connexin gene family reveal silencing of gjc1 (connexin45) by promoter hypermethylation in colorectal cancer, Epigenetics 6 (5) (2011) 602–609.

[9] E. OkochiTakada, K. Nakazawa, M. Wakabayashi, A. Mori, S. Ichimura, T. Yasugi, T. Ushijima, Silencing of the uchl1 gene in human colorectal and ovarian cancers, International Journal of Cancer 119 (6) (2006) 1338–1344.

[10] H. P. Joshi, I. V. Subramanian, E. K. Schnettler, G. Ghosh, R. Rupaimoole, C. Evans, M. Saluja, Y. Jing, I. Cristina, S. Roy, Y. Zeng, V. H. Shah, A. K. Sood, S. Ramakrishnan, Dynamin 2 along with microrna-199a reciprocally

regulate hypoxia-inducible factors and ovarian cancer metastasis, Proceedings of the National Academy of Sciences 111 (14) (2014) 5331–5336.

[11] C. Liu, S. Zhou, S. Begum, D. Sidransky, W. H. Westra, M. Brock, J. A. Califano, Increased expression and activity of repair genes tdp1 and xpf in non-small cell lung cancer, Lung Cancer 55 (3) (2007) 303–311.

[12] L. Papa, G. Manfredi, D. Germain, Sod1, an unexpected novel target for cancer therapy, in: Genes & cancer, 2014.

[13] C. M. Clements, R. S. McNally, B. J. Conti, T. W. Mak, J. P.-Y. Ting, Dj-1, a cancer- and parkinson's disease-associated protein, stabilizes the antioxidant transcriptional master regulator nrf2, Proceedings of the National Academy of Sciences 103 (41) (2006) 15091–15096.

[14] R. SaundersPullman, M. J. Barrett, K. M. Stanley, M. S. Luciano, V. Shanker, L. Severt, A. Hunt, D. Raymond, L. J. Ozelius, S. B. Bressman, Lrrk2 g2019s mutations are associated with an increased cancer risk in parkinson disease, Movement Disorders 25 (15) (2010) 2536–2541.

[15] Z.-C. Wang, Q. Gao, J.-Y. Shi, L.-X. Yang, J. Zhou, X.-Y. Wang, Y.-H. Shi, A.-W. Ke, G.-M. Shi, Z.-B. Ding, Z. Dai, S.-J. Qiu, J. Fan, Genetic polymorphism of the kinesin-like protein kif1b gene and the risk of hepatocellular carcinoma, PLOS ONE 8 (4) (2013) e62571.

[16] G. Li, M. Li, X. Liang, Z. Xiao, P. Zhang, M. Shao, F. Peng, Y. Chen, Y. Li, Z. Chen, Identifying dcn and hspd1 as potential biomarkers in colon cancer using 2d-lc-ms/ms combined with itraq technology, Journal of Cancer 8 (3) (2017) 479–489.

[17] Y. F. Han, G. W. Cao, Role of nuclear receptor nr4a2 in gastrointestinal inflammation and cancers, World Journal of Gastroenterology 18 (47) (2012) 6865–6873.

[18] W. J. Chia, B. L. Tang, Emerging roles for rab family gtpases in human cancer, Biochimica et Biophysica Acta (BBA) - Reviews on Cancer 1795 (2) (2009) 110–116.

[19] A. Korshunov, F. Sahm, D. Stichel, D. Schrimpf, M. Ryzhova, O. Zheludkova, A. Golanov, P. Lichter, D. T. W. Jones, A. von Deimling, S. M. Pfister, M. Kool, Molecular characterization of medulloblastomas with extensive nodularity (mben), Acta Neuropathologica (2018) 1–11.

[20] B. H. Lee, M. G. Taylor, P. Robinet, J. D. Smith, J. Schweitzer, E. Sehayek, S. M. Falzarano, C. Magi-Galluzzi, E. A. Klein, A. H. Ting, Dysregulation of cholesterol homeostasis in human prostate cancer through loss of abca1, Cancer Research 73 (3) (2013) 1211–1218.

[21] D. R. Robinson, Y. M. Wu, P. Vats, F. Su, R. J. Lonigro, X. Cao, S. Kalyana-Sundaram, R. Wang, Y. Ning, L. Hodges, A. Gursky, J. Siddiqui, S. A. Tomlins, S. Roychowdhury, K. J. Pienta, S. Y. Kim, J. S. Roberts, J. M. Rae, C. H. Van Poznak, D. F. Hayes, R. Chugh, L. P. Kunju, M. Talpaz, A. F. Schott, A. M. Chinnaiyan, Activating esr1 mutations in hormone-resistant metastatic breast cancer, Nature Genetics 45 (12) (2013) 1446–1451.

[22] Y. Chen, Y. Hu, H. Zhang, C. Peng, S. Li, Loss of the alox5 gene impairs leukemia stem cells and prevents chronic myeloid leukemia, Nature Genetics 41 (7) (2009) 783–792.

[23] D. S. Michaud, S. E. Daugherty, S. I. Berndt, E. A. Platz, M. Yeager, E. D. Crawford, A. Hsing, W.-Y. Huang, R. B. Hayes, Genetic polymorphisms of interleukin-1b (il-1b), il-6, il-8, and il-10 and risk of prostate cancer, Cancer Research 66 (8) (2006) 4525–4530.

[24] J. Iwashita, Y. Sato, H. Sugaya, N. Takahashi, H. Sasaki, T. Abe, mrna of muc2 is stimulated by il-4, il-13 or tnf-alpha through a mitogen-activated protein kinase pathway in human colon cancer cells, Immunology and cell biology 81 4 (2003) 275–82.

[25] C. Steidl, S. P. Shah, B. W. Woolcock, L. Rui, M. Kawahara, P. Farinha, N. A. Johnson, Y. Zhao, A. Telenius, S. B. Neriah, A. McPherson, B. Meissner, U. C. Okoye, A. Diepstra, A. Van Den Berg, M. Sun, G. Leung, S. J. Jones, J. M. Connors, D. G. Huntsman, K. J. Savage, L. M. Rimsza, D. E. Horsman, L. M. Staudt, U. Steidl, M. A. Marra, R. D. Gascoyne, Mhc class ii transactivator ciita is a recurrent gene fusion partner in lymphoid cancers, Nature 471 (7338) (2011) 377–383.

[26] M. Porcu, M. Kleppe, V. Gianfelici, E. Geerdens, K. De Keersmaecker, M. Tartaglia, R. Fo, J. Soulier, B. Cauwelier, A. Uyttebroeck, E. Macintyre, P. Vandenberghe, V. Asnafi, J. Cools, Mutation of the receptor tyrosine phosphatase ¡em¿ptprc¡/em¿ (cd45) in t-cell acute lymphoblastic leukemia, Blood 119 (19) (2012) 4476–4479.

[27] K. Okamura, T. Harada, S. Wang, K. Ijichi, K. Furuyama, T. Koga, T. Okamoto, K. Takayama, T. Yano, Y. Nakanishi, Expression of trkb and bdnf is associated with poor prognosis in non-small cell lung cancer, Lung Cancer 78 (1) (2012) 100–106.

[28] P. Vainio, L. Lehtinen, T. Mirtti, M. Hilvo, T. Seppnen-Laakso, J. Virtanen, A. Sankila, S. Nordling, J. Lundin, A. Rannikko, M. Orei, O. Kallioniemi, K. Iljin, Phospholipase pla2g7, associated with aggressive prostate cancer, promotes prostate cancer cell migration and invasion and is inhibited by statins, Oncotarget 2 (12) (2011) 1176–1190.

[29] J. S. Hale, B. Otvos, M. Sinyuk, A. G. Alvarado, M. Hitomi, K. Stoltz, Q. Wu, W. Flavahan, B. Levison, M. L. Johansen, D. Schmitt, J. M. Neltner, P. Huang, B. Ren, A. E. Sloan, R. L. Silverstein, C. L. Gladson, J. A. DiDonato, J. M. Brown, T. McIntyre, S. L. Hazen, C. Horbinski, J. N. Rich, J. D. Lathia, Cancer stem cellspecific scavenger receptor cd36 drives glioblastoma progression, STEM CELLS 32 (7) (2014) 1746–1758.