Supplementary Information for "Extracting the transitivity backbone of bipartite networks"

Lucía S. Ramírez,[1,2] Roya Aliakbarisani,[1,2] M. Ángeles Serrano,[1,2,3] and Marián Boguñá[1,2]

[1]*Departament de Física de la Matèria Condensada, Universitat de Barcelona, Martí i Franquès 1, E-08028 Barcelona, Spain*
[2]*Universitat de Barcelona Institute of Complex Systems (UBICS), Barcelona, Spain*
[3]*Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, E-08010 Barcelona, Spain*

CONTENTS

I. DESCRIPTION OF EMPIRIC DATASETS

Human metabolic network. The BiGG knowledgebase (*Biochemical, Genetic and Genomic*) [1] provides genome-scale metabolic network reconstructions using standardized nomenclature. BiGG integrates published metabolic models into a consistent framework, allowing comparisons across organisms and supporting export for external analysis. For *Homo sapiens*, the BiGG dataset comprises 2212 reactions involving 1224 metabolites (the original dataset in Ref. [2] reports 3743 reactions and 1509 metabolites).

YouTube group-membership network. The dataset [3] contains 37981 users and 12009 groups, with edges denoting memberships.

Plant–pollinator network. Robertson's historical compilation (1884–1916) documents flower-visiting insects in Macoupin County, Illinois, within a radius of approximately 16 km around Carlinville. His book [4] summarizes more than 15000 recorded visits by 1429 insect species to 456 flowering plant species, including detailed lists of bees, flies, beetles, butterflies, moths, and other visitors. These exhaustive records constitute one of the oldest and most complete bipartite datasets of plant–pollinator interactions, and have been widely used for ecological network analyses. Methods are detailed in Ref. [5].

World languages. The Unicode Common Locale Data Repository (CLDR) [6] provides estimates of the population that is literate in and comfortable using each language within each territory. This forms an affiliation network in which one set of nodes corresponds to territories and the other to languages, with edges indicating that a population in a given territory uses a given language. Specifically, CLDR's "Territory–Language Information" tables list, for each territory, the relevant languages, their share of the population, and literacy/writing percentages. The underlying data are partly derived from *Ethnologue* [7], but is supplemented and processed using other sources, including per-country census data, in order to include not only native but also people who are functional second-language.

II. BIPARTITE NETWORK WITH NOISE (SYNTHETIC BENCHMARKS)

We consider bipartite graphs with disjoint node sets A and B, where edges run only across sets. In our benchmarks, *noise* affects a controllable subset of B, while A represents the complementary node set. To stress-test the filter, we synthesize networks in which B mixes two link-formation mechanisms: (i) a *geometric* process that induces local similarity structure, and (ii) a *random* process that preserves degrees but removes geometric organization.

Mixture construction. Let N_A and N_B be the set sizes, with

$$B = B^{\mathbb{S}^1} \cup B^{\mathrm{CM}}, \qquad N_B = N_B^{\mathbb{S}^1} + N_B^{\mathrm{CM}}.$$

Nodes in $B^{\mathbb{S}^1}$ (signal) connect to A via the bipartite-$\mathbb{S}^1$ model [8–10]; nodes in B^{CM} (noise) connect via a (soft) configuration model with fixed expected degrees. Nodes in A and in $B^{\mathbb{S}^1}$ carry hidden variables (κ, θ), where κ sets expected degree and $\theta \in [0, 2\pi)$ is an angular similarity coordinate on the circle (latent similarity space). The signal-to-noise ratio

can be parameterized by

$$\mathcal{R} \equiv \frac{N_B^{\mathbb{S}^1}}{N_B^{\mathrm{CM}}},$$

or, equivalently, by the signal fraction $N_B^{\mathbb{S}^1}/N_B$.

A. Geometric links: the bipartite-$\mathbb{S}^1$ model

Edges between A and $B^{\mathbb{S}^1}$ are drawn independently with probability [8–10]

$$p_{\mathbb{S}^1}(\kappa_A, \kappa_B, \Delta\theta) = \left[1 + \left(\frac{R\,\Delta\theta}{\mu\,\kappa_A\kappa_B}\right)^{\beta}\right]^{-1}, \qquad (1)$$

where $\Delta\theta$ is the angular distance, $\beta > 1$ controls the sharpness of similarity effects (larger β yields stronger geometric clustering), $R = 2\pi N_A$ is the circle radius where nodes are homogeneously distributed, and $\mu = \frac{\beta}{2\pi\langle k_A^{\mathbb{S}^1}\rangle}\sin\left(\frac{\pi}{\beta}\right)$ fixes mean degrees. Denote by $\langle k_A^{\mathbb{S}^1}\rangle$ and $\langle k_B^{\mathbb{S}^1}\rangle$ the average degrees contributed by geometric links on sets A and $B^{\mathbb{S}^1}$, respectively; mass conservation implies

$$\langle k_A^{\mathbb{S}^1}\rangle = \frac{N_B^{\mathbb{S}^1}}{N_A}\,\langle k_B^{\mathbb{S}^1}\rangle. \qquad (2)$$

Intuitively, (1) favors connections between nodes with high expected degrees and small angular separation, producing recurrent neighborhoods and high bipartite clustering.

B. Random links: configuration model

Noise links between A and B^{CM} follow a soft configuration model:

$$p_{\mathrm{CM}}(\kappa_A, \kappa_B) = \frac{\mu_{\mathrm{CM}}\kappa_A\kappa_B}{1 + \mu_{\mathrm{CM}}\kappa_A\kappa_B}, \qquad (3)$$

with μ_{CM} chosen to match target mean degrees. Equivalently,

$$\mu_{\mathrm{CM}} = \frac{1}{N_B^{\mathrm{CM}}\langle k_B^{\mathrm{CM}}\rangle} = \frac{1}{N_A\langle k_A^{\mathrm{CM}}\rangle}, \qquad (4)$$

where $\langle k_B^{\mathrm{CM}}\rangle$ is the mean degree of CM nodes in B and $\langle k_A^{\mathrm{CM}}\rangle$ is their contribution to degrees in A. This process removes similarity structure while preserving expected degrees [11, 12].

C. Global accounting and degree distributions

After wiring both parts of B, the total number of edges satisfies

$$N_A\langle k_A\rangle = N_B^{\mathbb{S}^1}\langle k_B^{\mathbb{S}^1}\rangle + N_B^{\mathrm{CM}}\langle k_B^{\mathrm{CM}}\rangle, \qquad (5)$$

so the mean degree on A is determined by the mixture proportions and mean degrees on B. Unless stated otherwise, hidden degrees are sampled from power laws

$$\rho(\kappa) \propto \kappa^{-\gamma}, \quad \kappa \geq \kappa_0(\gamma, \langle k\rangle),$$

with $\gamma > 2$ and lower cutoffs ensuring finite means.

This construction yields ground-truth labels (geometric vs. random) for nodes in B, enabling quantitative evaluation of noise detection (e.g., via F-score) and assessment of downstream effects (percolation and node classification).

III. CLUSTERING COEFFICIENT FOR BIPARTITE NETWORKS

In unipartite graphs, the local clustering coefficient c_i measures the fraction of closed triplets around node i [13]. Triangles cannot exist in bipartite graphs, so bipartite analogues quantify closure via *squares* or via *two-step connections* through the opposite set [10, 14, 15]. High local closure is a hallmark of geometric organization in latent-space models [16, 17].

Degree-normalized bipartite clustering. We adopt a pairwise-closure definition that captures both the presence and the *strength* of local closure [10]:

$$c_{b,i} = \frac{2\,\tilde{T}_i}{k_i(k_i-1)}, \qquad \tilde{T}_i = \sum_{\alpha\neq\beta}\frac{T_{\alpha\beta}(i)}{\min(k_\alpha, k_\beta) - 1}. \qquad (6)$$

Here k_i is the degree of i; $T_{\alpha\beta}(i)$ is the number of common neighbors (in the opposite set) shared by the neighbor pair (α, β) of i; and the denominator $\min(k_\alpha, k_\beta) - 1$ is the maximum possible number of such shared neighbors for that pair (excluding i). Each pair thus contributes a fraction in $[0, 1]$, and $c_{b,i} \in [0, 1]$ averages these normalized contributions across all neighbor pairs. Compared with binary "at-least-one" closures [10], Equation (6) downweights pairs that close only weakly and upweights pairs that close through many common neighbors, improving sensitivity to geometric neighborhoods. Finally, $c_{b,i}$ is averaged over degree classes to define the bipartite clustering coefficient used in the filter.

Relation to alternative measures. Square-based coefficients [14, 15] count 4-cycles; motif-based and weighted extensions exist [12, 18]. We use (6) because it (i) is local and degree-aware, (ii) aligns with latent geometric models where proximity induces repeated co-neighborhoods, and (iii) yields degree-conditional null distributions under degree-preserving randomization, which we exploit for statistical filtering.

IV. RESOLUTION LIMIT

To further illustrate the degree-dependent resolution of the filter, Fig. S1 shows the cumulative fraction of nodes classified as noise as a function of degree for different parameter choices (from Figure 2 in the main text). Panel (c) and (d)

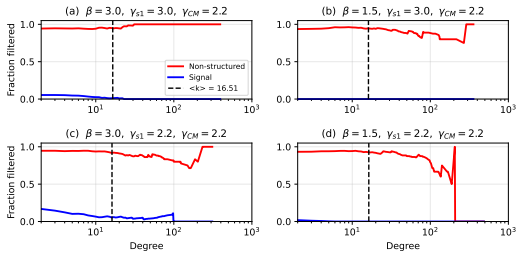


FIG. S1: Cumulative fraction of nodes classified as noise as a function of degree for each type of nodes. Noise nodes (red) are consistently identified and filtered across the full degree range, while signal nodes (blue) are preferentially preserved at large degrees. Signal tends to be misclassified at low degrees, where limited connectivity reduces the likelihood of forming local cycles. On the other hand, some high-degree nodes from the configuration model may be retained due to spuriously high clustering arising from their large number of connections. Cumulative fraction of nodes classified as noise as a function of degree for each type of nodes.

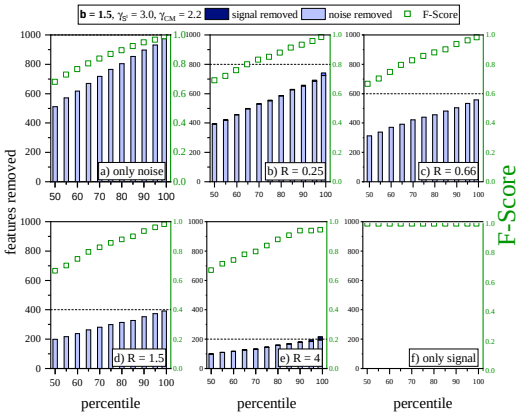


FIG. S2: **Performance of noise detection in synthetic bipartite networks** for $\beta = 1.5$, $\gamma_{\mathbb{S}^1} = 3.0$, $\gamma_{\mathrm{CM}} = 2.20$, and 1,000 nodes in set B, shown for different values of the signal-to-noise ratio R. For example, $R = 0.25$ corresponds to $N_{\mathbb{S}^1} = 200$ and $N_{\mathrm{CM}} = 800$. In all cases, the mean feature degree is $\langle k \rangle = 8$. Dashed lines indicate the number of noisy nodes in the set being filtered.

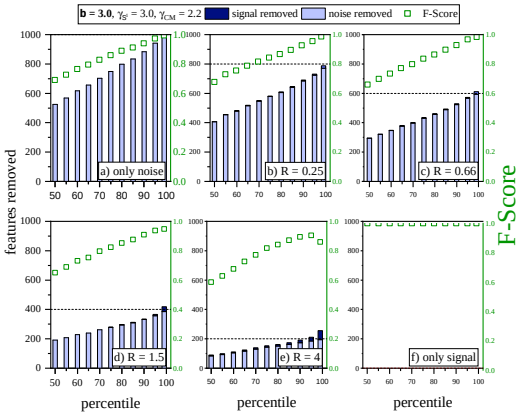


FIG. S3: **Performance of noise detection in synthetic bipartite networks** for $\beta = 3.0$, $\gamma_{\mathbb{S}^1} = 3.0$, $\gamma_{\mathrm{CM}} = 2.20$, and 1,000 nodes in set B, shown for different values of the signal-to-noise ratio R. For example, $R = 0.25$ corresponds to $N_{\mathbb{S}^1} = 200$ and $N_{\mathrm{CM}} = 800$. In all cases, the mean feature degree is $\langle k \rangle = 8$. Dashed lines indicate the number of noisy nodes in the set being filtered.

show extreme cases in which the signal is weakly geometrical making the classification more challenging.

For all cases, noise nodes are consistently identified and filtered over most of the degree range. Some hubs from the configurational model are retained in the backbone due to spurious high clustering due to the large number of connections. In contrast, signal nodes with low-degree are more prone to misclassifications due to limited connectivity and reduced cycle formation, high-degree nodes are reliably preserved. Nodes with larger degree provide more structural information, leading to a more robust identification of signal and noise.

V. RESULTS FOR DIFFERENT PARAMETERS

Figures S2, S3, and S4 show results of our filter method applied to synthetic bipartite networks with different parameters and unbalanced geometric and random sets.

VI. SCALING WITH SYSTEM SIZE

To assess how the method scales with system size, we analyze synthetic bipartite networks of increasing size, keeping the model parameters fixed. The results are shown in Fig. S5.

As the system size increases (from $N = 1000$ to $N = 20000$), the separation between signal and noise becomes progressively clearer in the clustering vs degree representation, specially for larger values of β. Nodes with random connections concentrate within the region defined by the null model, while nodes with geometric structure remain above it. At small sizes, fluctuations are more pronounced and the overlap between the two classes is larger, especially at low degrees. As N increases, these fluctuations are reduced and the bounds obtained from the rewired ensemble become smoother, improving the separation. This behavior is reflected in the classification performance. As shown in Fig. S6, the F-score increases with system size for all percentile thresholds and parameter choices. The improvement is more pronounced at higher percentiles, where the distinction between random and structured nodes becomes sharper.

These results are consistent with the fact that, in the configuration model, clustering goes to zero in the thermodynamic limit. As a consequence, the difference between nodes with purely random connections and those with geometric structure becomes more evident as the network grows, leading to a more accurate identification of the backbone.

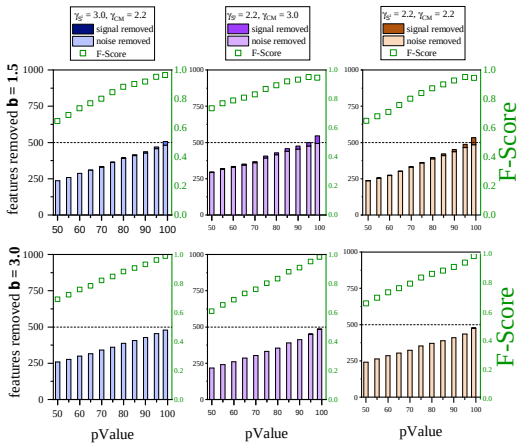
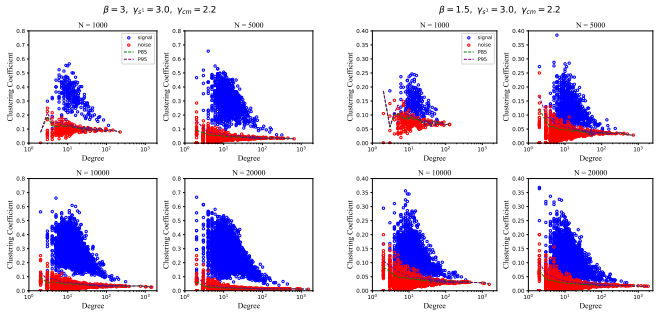


FIG. S4: **Performance of noise detection in synthetic bipartite networks**: The panels report the number of nodes classified as noise across different percentiles –corresponding to different significance levels p. Light bars denote true positives –noisy nodes correctly identified–whereas dark bars denote false positives –signal nodes misclassified as noise. Each row corresponds to a different value of β (indicated in the figure), and each column represents combinations of $\gamma_{\mathbb{S}^1}$ and γ_{CM} for the degree distributions of sets $B^{\mathbb{S}^1}$ and B^{CM}. In all cases, the number of nodes of each type is equal to 500, $\langle k_{\mathrm{A}} \rangle = 16$, $\langle k_{\mathrm{B}^{\mathbb{S}^1}} \rangle = 10$, and $\langle k_{\mathrm{B}^{CM}} \rangle = 6$. Green squares show the F-Score at each percentile, capturing the precision-recall trade-off. For percentiles above 0.85%, F-Score values are typically $\gtrsim 0.9$, demonstrating the method's robustness.



(a) $\beta = 3$. (b) $\beta = 1.5$.

FIG. S5: Clustering coefficient vs. degree for different system sizes ($N = 1000, 5000, 10000, 20000$) and model parameters ($\gamma_{S^1} = 3.0$, $\gamma_{CM} = 2.2$). Signal (blue) and noise (red) nodes are shown together with percentile bounds ($p = 85, 95$) obtained from the rewired ensemble. As the system size increases, the separation between signal and noise becomes clearer.

VII. PERFORMANCE FOR INCREASING AVERAGE DEGREE

To further assess the robustness of the method, we analyze its performance as the average degree of the network increases within the sparse regime. Increasing the mean degree enhances the local connectivity of nodes and leads to a higher number of potential cycles. However, under the degree-preserving configuration model, the expected clustering remains smaller even as the mean degree grows. As a result, the contrast between nodes whose connectivity is compatible with the null model and those exhibiting structured organization is not lost.

To illustrate this behavior, we performed additional experiments on synthetic networks with progressively larger mean degree, up to $\langle k \rangle \sim 80$. The results, reported in Figs. S7 and S8, show that the method remains robust across this range. In particular, nodes associated with structured connectivity continue to exhibit clustering values significantly above the random baseline, while nodes compatible with the null model remain confined within the expected region. These results indicate that, within the sparse regime, increasing the mean degree does not degrade the discriminative power of the method.

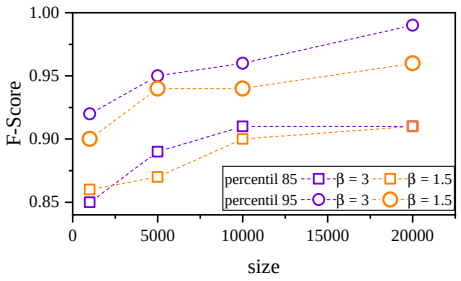
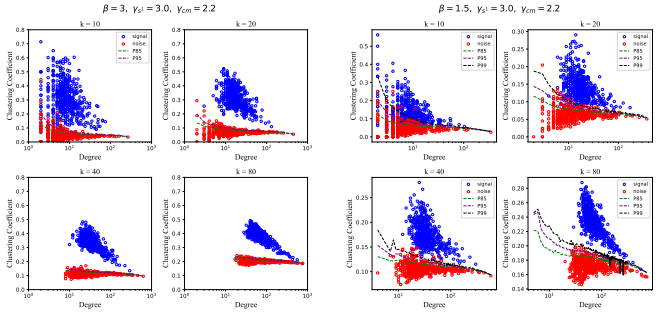


FIG. S6: F-score as a function of system size for different percentile thresholds for $\gamma_{S^1} = 3.0$, $\gamma_{CM} = 2.2$ different values of β . The classification accuracy improves with system size.



(a) $\beta = 3$. (b) $\beta = 1.5$.

FIG. S7: Clustering coefficient vs. degree for different mean degree. Signal (blue) and noise (red) nodes are shown together with percentile bounds ($p = 85, 95$) obtained from the rewired ensemble. As the mean degree of the network increases, the separation between signal and noise becomes clearer.

VIII. VALIDATION OF THE INFERRED β PARAMETER

We consider synthetic bipartite networks composed of a geometric component and a fraction of noisy nodes. The ground-truth value of β is the one used to generate the con-

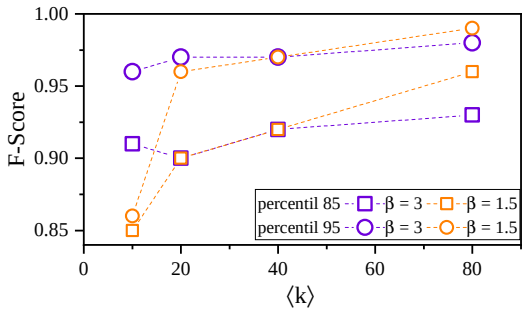


FIG. S8: F-score as a function of system size for different percentile thresholds for $\gamma_{S^1} = 3.0$, $\gamma_{CM} = 2.2$ different values of β as the mean degree of the network grows. The classification accuracy improves with system size.

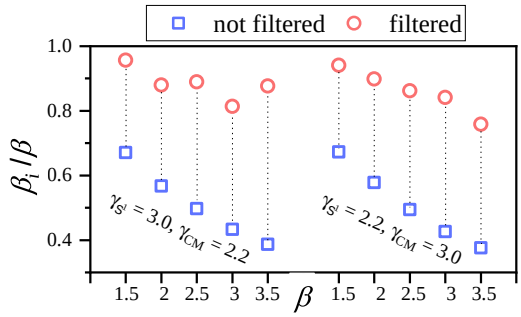


FIG. S9: Ratio between the inferred β_i and the actual β for different network configurations. Blue squares represent the unfiltered network, while pink circles correspond to the filtered network (percentile 95th). The dotted lines highlight the improvement in β_i after filtering. After filtering, the inferred β_i approaches the true β, demonstrating the effectiveness of the filtering process.

nections of nodes in the $B^{\mathbb{S}^1}$ set. For each realization, we estimate β_i both before and after applying the filtering procedure.

As shown in Fig. S9, when the full network (including both geometric and noisy nodes) is considered, the inferred parameter β_i is systematically lower than the ground-truth value β. After filtering, β_i shifts closer to β, despite the presence of residual noise.

This effect is summarized by the ratio β_i/β, reported in Supplementary Fig. S9 for different network configurations.

IX. EFFECT OF SWAPPING THE BIPARTITE NODE SETS

Consider a bipartite network composed of two node sets, A and B. We apply the filtering procedure to nodes in set A, while nodes in set B consist of a mixture of signal and noise. The proposed filtering procedure is applied to set B of nodes in a bipartite network, under the assumption that this set may contain noisy elements. In many cases, the choice of what set to filter arises naturally from the interpretation of the data. For instance, in the case of bipartite networks between nodes and their features, as used in machine learning applications, the nodes to be classified are obviously the type-A nodes, while their features are the type-B nodes. This is because one can always assign features to nodes that are completely unrelated to the similarity space one aims to uncover. Consider, for example, the Cora dataset. It is a bipartite network between scientific articles (type-A nodes) and their features, defined as a bag of words extracted from the articles' abstracts (type-B nodes). Since articles focus on particular scientific fields and subfields, we expect node classification to be possible by examining how similarly related articles use similar words. This is the key idea behind machine learning techniques for classifying articles into categories. Indeed, many of these words or features form specific vocabularies associated with highly specialized scientific fields, and therefore they carry a great deal of structural information. However, many words are generic and may or may not appear in the abstract of any given article. These are therefore non-informative and can be regarded as noisy type-B nodes following a heterogeneous degree distribution, most likely arising from Zipf's law, and connecting to type-A nodes at random.

In other cases, however, the choice of which node set to filter is less clear. To assess the effect of this choice, we analyze the outcome of applying the filter under a non-natural assignment of node types.

Consider a bipartite network in which type-A nodes are noise-free, whereas a fraction f_{rand} of type-B nodes are random and the remaining fraction are non-random. If we interchange the roles of sets A and B and apply the filter to this reversed representation, an asymmetry emerges. The original type-A nodes are connected to both random and non-random type-B nodes, whereas non-random type-B nodes are connected only to non-random type-A nodes. Consequently, the clustering coefficient of type-A nodes acquires a random component whose strength depends on the fraction of noisy type-B nodes.

When f_{rand} is very small, the dominant contribution to the clustering of type-A nodes still arises from non-noisy type-B nodes. Their clustering coefficient therefore remains high and independent of system size, and the filter removes almost no type-A nodes. In the opposite limit, when $f_{rand} \approx 1$, the clustering of type-A nodes is determined almost entirely by random connections. In this case, type-A nodes also behave as random nodes, and the filter removes nearly all of them.

The intermediate regime is the most informative. For a moderate fraction of noisy nodes, $f_{rand} \approx 0.5$, approximately half of the links incident on type-A nodes come from noisy type-B nodes, thereby reducing their clustering coefficient. This effect is particularly strong for low-degree nodes, but less pronounced for high-degree ones. Random fluctuations make low-degree type-A nodes more likely to have a large fraction of their neighbors among noisy type-B nodes, which suppresses their clustering coefficient and leads the filter to classify them as noisy. High-degree type-A nodes, by contrast, retain enough connections to non-noisy type-B nodes for their clustering coefficient to remain relatively large.

Therefore, in this regime the filter removes mainly a fraction of low-degree type-A nodes while preserving high-degree

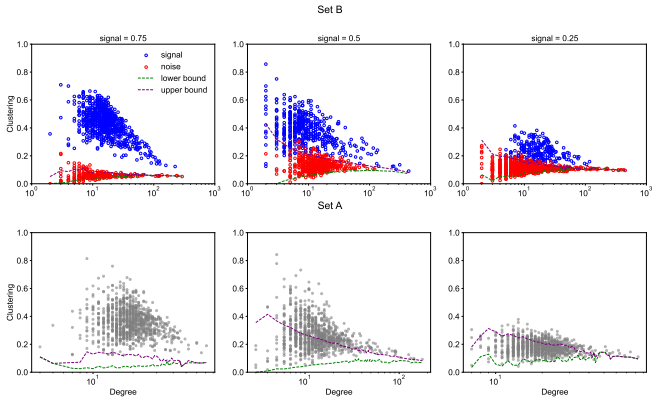


FIG. S10: Clustering coefficient as a function of degree for a networks with $\beta = 3$, $\gamma_{S^1} = 3$, and $\gamma_{CM} = 2.2$. **Top row (Set B):** Natural assignment of Set B, where signal (blue) and noise (red) nodes are well identified. Signal nodes exhibit systematically higher clustering, and the percentile bounds (dashed lines) capture the distinction across the degree range. **Bottom row (Set A):** Result after swapping the roles of the two node sets.

nodes in the backbone. This reflects the fact that nodes whose connectivity is strongly affected by random links lose a clear structural interpretation, whereas nodes with more stable connectivity patterns are retained.

To illustrate this effect, we apply the filtering procedure to both layers of the bipartite network for three networks with different signal-to-noise ratios. In all cases, $\beta = 3$, $\gamma_{S^1} = 3$, and $\gamma_{CM} = 2.2$. The results are shown in Fig. S10.

In Fig. S10, we report both the upper and lower bounds corresponding to the 95th percentile obtained from the rewired ensemble. The upper bound is used in the main text to define the filtering threshold, while the lower bound corresponds to the minimum clustering expected under the same random baseline.

The top row (Set B) shows the clustering profile when filtering the layer that contains both geometrically structured and noisy connections. The bottom row (Set A) shows the outcome after swapping the roles of the two node sets.

In Set B, the lower bound closely follows the lower edge of the data distribution, and together with the upper bound delimits the region where noisy nodes are expected to lie across the degree range. In contrast, when the roles of the node sets are exchanged (Set A), the lower bound lies systematically below the bulk of the data, with the separation becoming more pronounced as the level of noise in Set B decreases. This reflects the fact that nodes in Set A exhibit higher clustering than expected for purely random connections, as they retain links to geometrically structured neighbors.

X. FEATURE SELECTION TECHNIQUES

We compare our filter with common feature-selection methods used in machine learning. Feature selection reduces dimensionality by retaining variables most relevant to a prediction task, which can shorten training time, mitigate overfitting, and improve interpretability [19, 20]. This is particularly valuable in high-dimensional settings where many features are redundant or uninformative.

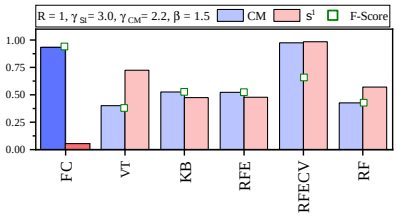


FIG. S11: Comparison of the fraction of features detected as noise (blue) and signal mistakenly flagged as noise (pink) across methods in `sklearn.feature_selection`: Variance Threshold (VT), SelectKBest (KB), Recursive Feature Elimination (RFE), RFE with cross-validation (RFECV), and Random Forest (RF). FC denotes our clustering-based filter. In the synthetic bipartite benchmarks, each node set has size 1000 and the filtered mixture has ratio $R = 1$ (500 geometric and 500 non-geometric features).

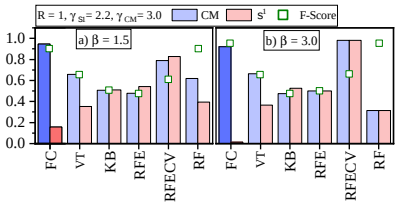


FIG. S12: The same as in Fig. S11 for different model parameters β, $\gamma_{\mathbb{S}1}$, and γ_{CM}.

We benchmarked against methods from `sklearn.feature_selection` [21]: Variance Threshold (VT), SelectKBest (KB), Recursive Feature Elimination (RFE), RFE with cross-validation (RFECV), and Random Forest (RF; importance-based). Results are shown in Figs. S11 and S12 for different parameters of the synthetic networks. For VT, KB, RFE, and RF, the fraction of features flagged as noise is close to the random baseline –roughly half of each class– indicating limited discrimination in our setting. As expected for sparse data, RFECV prunes nearly all features. For all methods except RFECV, features identified as noise are predominantly low-degree. A key limitation is that univariate filters (VT, KB) assess features independently and ignore interactions, while wrapper/embedded methods (RFE, RF) are not tailored to the probabilistic, bipartite structure of our data. Instead, our clustering-based method (FC) applied to the same data is able to detect more than 90% of noisy features while producing a small fraction of false positives, leading to an F-score above 90%.

XI. NEURAL NETWORK MODELS

This section provides details of two well-known neural network models used in the main paper for node classification

tasks. The models are fed with the original, filtered, and randomly filtered feature matrices to evaluate the impact of our statistical filter on node classification accuracy.

MLP. It is a vanilla neural network that takes node feature vectors as input, undergoing a sequence of linear transformations optionally followed by non-linear activation functions. This process enables the model to learn novel embeddings for nodes within a Euclidean space.

HNN [22]. It extends standard neural network operations to the Poincaré model of hyperbolic geometry, functioning as a specialized variant of the MLP in hyperbolic space. By leveraging the unique properties of hyperbolic geometry, HNNs offer an enriched framework for representation learning, particularly effective at capturing complex patterns and hierarchical structures.

Node labels are predicted by applying a softmax to the final layer outputs and selecting the class with the highest probability.

A. Model Configurations

In all experiments, nodes are divided into training (70%), validation (15%), and test (15%) sets. Both models are trained using the Adam optimizer [23] with a learning rate of 10^{-2}, dropout 0.2, and weight decay 0.001. The models include 2 hidden layers of dimention 16. Reported accuracies represent averages over 100 independent runs with different random seeds. Table I summarizes the hyperparameters of the models.

TABLE I: Hyperparameters of the MLP and HNN models.

Model	Layers	Dimensions	Optimizer	Activation	Dropout	Weight Decay
MLP	2	16	Adam	-	0.2	0.001
HNN	2	16	Adam	-	0.2	0.001

[1] J. Schellenberger, J. Park, T. M. Conrad, and B. Ø. Palsson, Bigg: a biochemical genetic and genomic knowledgebase of large-scale metabolic network reconstructions, BMC Bioinformatics **11**, 213 (2010).

[2] N. D. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. O. Palsson, Global reconstruction of the human metabolic network based on genomic and bibliomic data, Proceedings of the National Academy of Sciences **104**, 1777 (2007).

[3] R. A. Rossi and N. K. Ahmed, The network data repository with interactive graph analytics and visualization, in *AAAI* (2015).

[4] C. Robertson, Flowers and insects: lists of visitors to four hundred and fifty-three flowers (1929), carlinville, IL, USA. See `http://www.ecologia.ib.usp.br/iwdb/html/robertson_1929.html`.

[5] J. C. Marlin and W. E. LaBerge, The native bee fauna of carlinville, illinois, revisited after 75 years: a case for persistence, Conservation Ecology **5** (2001).

[6] U. Consortium, Unicode common locale data repository (cldr) (2024), available at: `https://github.com/unicode-org/cldr` (Accessed: 2024-11-15).

[7] D. M. Eberhard, G. F. Simons, and C. D. Fenning, Ethnologue: Languages of the world (2015).

[8] M. Á. Serrano, M. Boguñá, and F. Sagués, Uncovering the hidden geometry behind metabolic networks, Mol. BioSyst. **8**, 843 (2012).

[9] M. Kitsak, F. Papadopoulos, and D. Krioukov, Latent geometry of bipartite networks, Physical Review E **95**, 032309 (2017).

[10] R. Aliakbarisani, M. Á. Serrano, and M. Boguñá, Feature-enriched hyperbolic network geometry, Physical Review Research **7**, 033036 (2025).

[11] M. E. J. Newman, The structure and function of complex networks, SIAM Review **45**, 167 (2003).

[12] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, The architecture of complex weighted networks, Proceedings of the National Academy of Sciences **101**, 3747 (2004).

[13] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, Nature **393**, 440 (1998).

[14] P. G. Lind, M. C. González, and H. J. Herrmann, Cycles and clustering in bipartite networks, Physical Review E **72**, 056127 (2005).

[15] P. Zhang, J. Wang, X. Li, Z. Di, and Y. Fan, Clustering coefficient and community structure of bipartite networks, Physica A: Statistical Mechanics and its Applications **387**, 6869 (2008).

[16] M. Á. Serrano and M. Boguñá, *The Shortest Path to Network Geometry: A Practical Guide to Basic Models and Applications*, Elements in Structure and Dynamics of Complex Networks (Cambridge University Press, 2022).

[17] D. Krioukov, Clustering implies geometry in networks, Physical Review Letters **116**, 208302 (2016).

[18] T. Opsahl and P. Panzarasa, Clustering in weighted networks, Social Networks **31**, 155 (2009).

[19] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research **3**, 1157 (2003).

[20] Y. Saeys, I. Inza, and P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics **23**, 2507 (2007).

[21] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, Scikit-learn: Machine learning in python, Journal of Machine Learning Research **12**, 2825 (2011).

[22] O. Ganea, G. Becigneul, and T. Hofmann, Hyperbolic neural networks, in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., Red Hook, NY, USA, 2018) pp. 5345–5355.

[23] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980 (2014).