Nonlinear Science 8 (2026) 100155

Contents lists available at ScienceDirect

Nonlinear Science

journal homepage: www.elsevier.com/locate/nls

Research paper

# Higher-order architecture of metastatic networks: A statistical mechanics approach via validated hypergraphs

David H. Margarit [a,b,*], Marcela V. Reale [a,b], Anthony C.C. Coolen [c]

[a] Instituto de Ciencias (ICI), Universidad Nacional de General Sarmiento (UNGS), J.M. Gutierrez 1150, Los Polvorines, B1613, Buenos Aires, Argentina

[b] Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Godoy Cruz 2290, Ciudad Autónoma de Buenos Aires, C1425, Buenos Aires, Argentina

[c] Donders Centre for Neuroscience (DCN), Faculty of Science, Radboud University, Heyendaalseweg 135, Nijmegen, 6525 AJ, The Netherlands

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.17542956

*Keywords:*
Hypergraphs
Complex networks
Statistical mechanics
Metastatic organotropism
Null models
Nonlinear dynamics in biological systems

ABSTRACT

Metastasis is a complex, stochastic systemic process characterised by non-random spread patterns to secondary organs. Conventional dyadic projections fragment multi-organ interactions into independent events, effectively discarding irreducible higher-order topological information and the synergistic nature of clonal dissemination. This limitation often leads to conflating first-order topological properties with intrinsic biological affinities, particularly in the absence of validation against structure-aware null models.

To address this methodological gap, we present a higher-order hypergraph framework that preserves simultaneous colonisation patterns and accounts for emergent architectures that cannot be reduced to simple pairwise interactions. Our approach incorporates: (1) a functional partition of the network nodes $(V_S, V_T)$ to distinguish organs with documented reseeding capacity; and (2) a directed association matrix $\mathcal{S}$ quantifying the non-linear information flux and dependencies among source organs.

We implement a rigorous validation protocol by exploring the constrained null-space of the metastatic hypergraph using a Metropolis–Hastings MCMC algorithm. By contrasting the empirical data against a degree-preserving Canonical Ensemble (Configuration Model), this framework acts as a structural filter, demonstrating that many observed metastatic hotspots are spurious correlations emergent from nodal degree heterogeneity rather than genuine biological co-selection. By de-convoluting effective affinities from degree-driven effects, our method identifies a robust structural backbone that remains invariant under topological perturbations. This work shifts the focus from counting isolated events to quantifying the statistical mechanics of non-linear disease progression in complex biological networks.

## 1. Introduction

Metastasis represents a quintessential nonlinear complex process [1] where uncontrolled cell proliferation leads to a non-random, systemic dissemination to distant organs [2,3]. This process exhibits organotropism, emerging from a high-dimensional interplay of anatomical, physiological, and molecular factors [4–6]. Characterising the global architecture of these dissemination pathways remains a central challenge in understanding the collective behaviour and emergent properties of metastatic systems [7].

* Corresponding author at: Instituto de Ciencias (ICI), Universidad Nacional de General Sarmiento (UNGS), J.M. Gutierrez 1150, Los Polvorines, B1613, Buenos Aires, Argentina.

*E-mail addresses:* dmargarit@campus.ungs.edu.ar (D.H. Margarit), mreale@campus.ungs.edu.ar (M.V. Reale).

https://doi.org/10.1016/j.nls.2026.100155
Received 9 April 2026; Received in revised form 26 May 2026; Accepted 29 May 2026
Available online 3 June 2026
3050-5178/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

The clinical foundation of this work integrates organotropism data from six systematic reviews: Fares et al. (2020) [8] on molecular principles; Gao et al. (2019) [9] on "congenial soil"; Agrawal et al. (2025) [10] on mechanical signatures; Wang et al. (2021) [11] on metabolic adaptation; Welch and Hurst (2019) [12] on core hallmarks; and Bautista and López-Cortés (2025) [13] on therapeutic targets. This corpus is supplemented by the Database of Metastatic Patterns of Spread [14], providing quantitative statistics, and the National Cancer Institute [15] for clinical standards.

To ensure robustness and satisfy the need for technically correct data-driven modelling, an association was included in the incidence matrix only if validated by at least 6 out of 8 ($\geq$75%) independent clinical sources, thereby filtering out anecdotal reports and prioritising robust, reproducible metastatic patterns. This threshold aims to minimise false positives by retaining only consistently documented patterns, at the cost of potentially excluding rarer but real associations.

Network science provides a powerful framework to model this complexity, framing metastasis as a system of stochastic dependencies between organs [16–19]. Whilst we utilise concepts such as "information flow" to describe the systemic architecture, it is essential to clarify that our framework characterises static structural co-occurrence rather than temporal transition paths. In this context, "flow" refers to the statistical influence that the colonisation of one organ exerts on the probability of observing involvement in another within the same metastatic profile.

Recently, Gosak et al. [20] comprehensively examined the use of network methodologies for the morphological and structural quantification of biological systems, compiling evidence ranging from cell packing in epithelia to the architecture of brain connectomes and anatomical networks. This body of work corroborates that modelling structural connectivity as a network enables the quantitative characterisation of tissue organisation and function across multiple scales. However, conventional topological representations are restricted to pairwise interactions [21], which limits their applicability to systems governed by collective dynamics. In neuroscience, higher-order interactions (e.g., the collective synchronisation of three or more neurons) have been shown to induce complex spatiotemporal phenomena, such as chimera states, which cannot be explained within the pairwise connectivity paradigm [22]. Additionally, the temporal evolution of these higher-order properties plays a crucial role in modulating neuronal synchronisation [23]. These findings underscore a general principle: in complex biological systems, group interactions encode irreducible information that dyadic projections fail to capture.

Despite its potential, current network models face two key limitations in capturing nonlinear dependencies. First, conventional graphs are restricted to pairwise interactions, artificially decomposing higher-order synergistic patterns into isolated binary links and causing a structural fragmentation that obscures the multi-organ dissemination architecture. Second, many models lack rigorous validation against architecture-aware null models, making it difficult to distinguish genuine biological signal from stochastic noise or varying nodal degree centrality (i.e., the heterogeneous frequency with which different primary tumours metastasise to each site).

Consequently, a unified framework capable of capturing the higher-order architecture of metastatic dissemination is still lacking. Addressing this methodological gap requires a data-driven framework that simultaneously preserves multi-site dependencies, incorporates biologically meaningful functional distinctions between nodes, and validates observed patterns against appropriately constrained null models derived from statistical mechanics.

Current models often conflate organ "popularity", the inherent frequency of colonisation, with genuine biological affinity. There is a critical need for a framework that de-convolutes these factors. To address these challenges, we develop a higher-order framework with two main components. First, we employ a hypergraph representation [24–26] that preserves the inherent architecture of multi-organ patterns, treating metastasis as a correlated multi-body system. Second, we introduce a hierarchical statistical validation approach that tests associations against a degree-preserving configuration model using a Metropolis–Hastings MCMC protocol [27], explicitly controlling for organ-specific metastatic susceptibility.

Translating concepts from network biology to clinical reality requires addressing challenges such as organ degree centrality bias. Although hypergraph models have proven useful for studying molecular interactions, such as cancer stem cell marker networks [28,29], the clinical architecture introduces specific confounds. Our framework builds upon statistical safeguards [9] to help ensure that the higher-order patterns identified are not merely spurious correlations of heterogeneous organ involvement.

Primary cancer types were selected based on global epidemiological relevance and the need to capture the full spectrum of clinically distinct organotropisms [30–32]. Our dataset (Table 1) includes both prevalent cancers and a broader range of solid tumours (e.g., melanoma, sarcomas), encompassing haematogenous, lymphatic, and transcoelomic dissemination routes. This is critical because our approach identifies the topological signature of common clonal origins that dyadic representations typically discard.

We define metastatic sites broadly to include both distant and regional targets, such as lymph nodes and the peritoneum. The term "organ" is used anatomically to include solid organs and key tissues recognised as critical nodes in the nonlinear spread. This study is restricted to solid tumour malignancies; haematopoietic cancers follow a different dynamical trajectory and fall outside the scope of our hypergraph model [33].

Our analytical approach comprises three stages: (1) integrating organotropic patterns to define higher-order co-occurrence events; (2) representing these data as a hypergraph to preserve multi-site dependencies; and (3) introducing a functional partition criterion ($V_S$ and $V_T$) that de-convolutes organs by their functional role. Probabilistic associations are tested against an Erdős–Rényi analogue and a degree-preserving configuration model, identifying co-occurrence patterns that cannot be explained by random chance.

Although metastatic progression is inherently dynamic, characterising its underlying global architecture is a prerequisite for mechanistic modelling. Our static hypergraph framework focuses on this structural description, identifying higher-order constraints that may inform future stochastic or kinetic models. Collectively, these components provide a formal method to characterise the emergent architecture of metastatic spread.

**Table 1**
Classification of metastatic patterns utilised in the hypergraph construction. Target organs are categorised by reported clinical affinity (Common, Occasional, or Rare) based on established clinical consensus. This stratification serves as a validation baseline for our statistical mechanics framework; by grouping organs according to observed epidemiological frequency, we can assess whether the hypergraph backbone successfully recovers known high-affinity metastatic relationships whilst distinguishing them from noise driven by degree heterogeneity.

| Primary sites | Potential recipient organs for metastasis | | |
|---|---|---|---|
| | Common | Occasional | Rare |
| Head & neck | Lymph nodes | Lung, bone, liver | Brain, skin |
| Lung | Bone, brain, Lymph nodes, liver, adrenal | Pleura | Kidney, thyroid, diaphragm, peritoneum, skin, pancreas, spleen, breast |
| Kidney | Lung | Liver, bone | Brain, adrenal, skin, thyroid, pancreas, spleen, breast |
| Pancreatic | Liver, peritoneum | Lung, bone | Stomach, colon, brain |
| Bladder | Lung, bone, liver | Lymph nodes, peritoneum, adrenal | Kidney, rectum, colon, prostate, ureter, pleura, brain |
| Testicular | Lymph nodes, lung | Liver | Bone, brain |
| Melanoma | Lymph nodes, lung, liver, brain, bone | Peritoneum | Adrenal, pancreas, spleen |
| Uterine | Lung | Lymph nodes, liver, ovary | Bone, peritoneum |
| Rectal | Liver, lung | Peritoneum, bone | Brain, adrenal, ocular |
| Prostate | Bone, Lymph nodes | Lung, liver | Peritoneum, brain, adrenal, kidney, pancreas |
| Thyroid | Lung, bone | Lymph nodes | Brain, liver, peritoneum |
| Oesophageal | Lymph nodes, liver, lung | Bone | Brain, adrenal |
| Breast | Lymph nodes, bone, lung, liver | Pleura, brain, adrenal, skin | Peritoneum, pancreas, spleen, ocular |
| Stomach | Liver, peritoneum | Lung, bone, Lymph nodes | Ovary |
| Colon | Liver | Lung, peritoneum | Bone, brain, bladder, stomach, skin, ocular, spleen |
| Ovarian | Peritoneum, diaphragm, liver | Lymph nodes, pleura | Bone, brain, colon, lung, skin |
| Liver | Lung, bone, Lymph nodes, peritoneum | Adrenal | Brain |
| Cervical | Lymph nodes, lung | Bone, liver, peritoneum | Brain |
| Soft tissue sarcoma | Lung | Bone, liver | Brain, adrenal |
| Ocular | Liver | Lung, bone | Skin, brain |
| Bone sarcoma | Lung | Liver | Brain, kidney, adrenal |

## 2. Methodology: A higher-order framework for metastatic mapping

The metastatic incidence matrix underlying our model was constructed by synthesising organotropic patterns from six systematic reviews [8–13] and two curated clinical databases [14,15]. From a systems-modelling perspective, primary sites were aggregated into 21 representative categories to ensure a discrete state space while maintaining statistical robustness. Specifically, broad clinical labels such as "kidney cancer" or "soft tissue sarcoma" were treated as aggregate nodes encompassing their respective histological subtypes [34].

This data-integration protocol (detailed in Table 1) allows the hypergraph representation to account for both ubiquitous systemic pathways and high-specificity organotropic niches. By integrating multiple established sources, we mitigate potential biases and ensure that the model input reflects a conserved structural core rather than institutional variability [35].

### *2.1. From dyadic interactions to hypergraph architectures*

Metastatic dissemination is characterised by collective patterns of colonisation that frequently involve multiple organs simultaneously [28]. These co-occurrences represent higher-order interactions where a single primary tumour (the driver) induces a specific multi-site configuration in the secondary organ space.

In a traditional network $G = (V, E')$, each edge represents a binary projection of the metastatic process. Such dyadic decomposition fragments a single multi-organ pattern into independent events and discards the topological information regarding their common origin. This approach implicitly assumes that metastatic spread to multiple sites is a collection of independent events, an assumption that fails to capture the nonlinear dependencies where the co-occurrence of metastases in a set of organs $\{O_i, O_j, O_k\}$ is not merely the sum of independent dyadic probabilities.

To preserve this architecture, we model the system as a hypergraph $H = (V, E)$, where hyperedges can connect any number of vertices, capturing non-additive interactions [36]. Each hyperedge $e_j \subseteq V$ represents the ensemble of organs reported as metastatic targets for the $j$th primary site. Our model consists of:

- **The set of vertices (nodes),** $V = \{v_1, v_2, \dots, v_M\}$, corresponding to the $M = 23$ organs acting as metastatic targets:

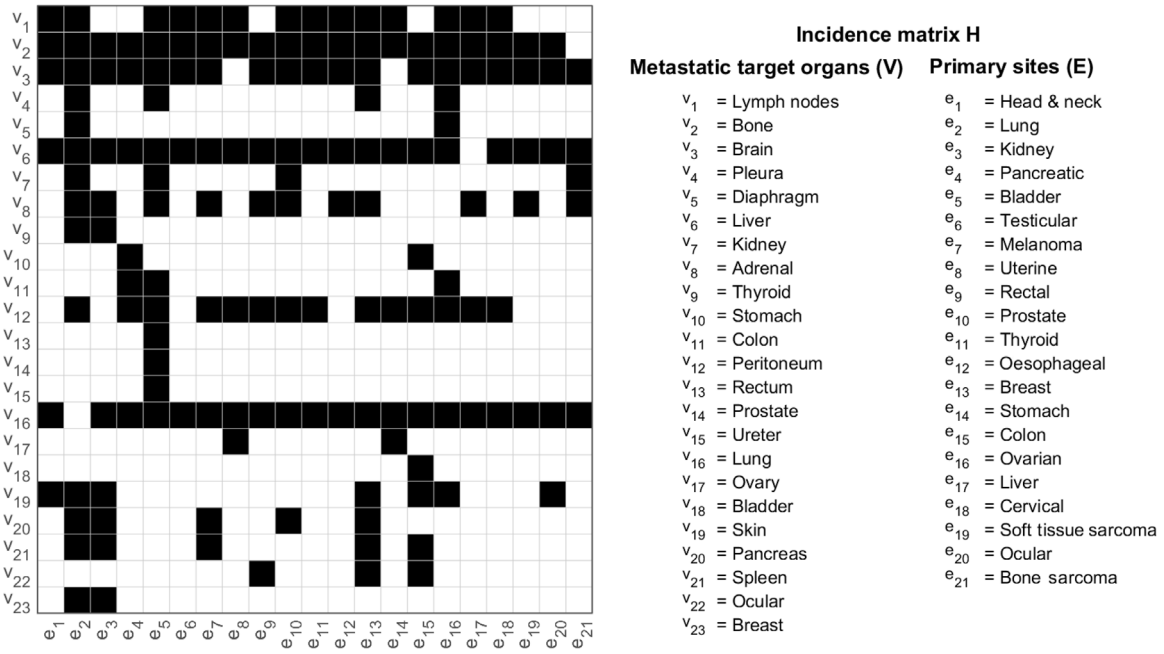



**Fig. 1.** Visualisation of the incidence matrix $H$ (Eq. (1)), which maps the metastatic landscape across the patient cohort. Rows correspond to metastatic target organs (vertices $V$), and columns represent primary cancer sites (hyperedges $E$). Black squares indicate documented associations ($H_{ij} = 1$), while white areas signify no reported metastatic events. This sparse structure illustrates the non-random, heterogeneous distribution of metastatic organotropism, serving as the foundational input for our hypergraph-based statistical mechanics framework.

$V$ = {Lymph nodes, Bone, Brain, Pleura, Diaphragm, Liver, Kidney, Adrenal, Thyroid, Stomach, Colon, Peritoneum, Rectum, Prostate, Ureter, Lung, Ovary, Bladder, Skin, Pancreas, Spleen, Ocular, Breast}.

- **The set of hyperedges**, $E = \{e_1, e_2, \dots, e_N\}$, corresponding to the $N = 21$ primary cancers acting as the system drivers:

$E$ = {Head and Neck, Lung, Kidney, Pancreatic, Bladder, Testicular, Melanoma, Uterine, Rectal, Prostate, Thyroid, Oesophageal, Breast, Stomach, Colon, Ovarian, Liver, Cervical, Soft tissue sarcoma, Ocular, Bone sarcoma}.

A metastatic association was encoded as $H_{ij} = 1$ only when validated by at least 75% of clinical sources (Table 1). This binarisation acts as a high-pass filter, isolating the invariant structural core of the network. Fig. 1 visualises the incidence matrix $H \in \{0,1\}^{M\times N}$, defined as:

$$H_{ij} = \begin{cases} 1 & \text{if node } v_i \text{ is a documented target for the hyperedge } e_j, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

### 2.2. Heterogeneity and degree distributions

In hypergraph theory, the structural role of nodes and edges is characterised by their degrees. The diagonal degree matrices $D_v$ and $D_e$ provide a quantitative measure of the system's heterogeneity:

$$D_v = \text{diag}(14, 20, 19, 4, 2, 20, 4, 11, 2, 2, 3, 14, 1, 1, 1, 20, 2, 1, 7, 5, 5, 3, 2) \tag{2}$$

and

$$D_e = \text{diag}(6, 14, 10, 7, 13, 5, 9, 6, 7, 9, 6, 6, 12, 6, 10, 10, 6, 6, 5, 5, 5). \tag{3}$$

The vertex degrees ($D_v$) reveal the presence of global attractors (hubs) within the metastatic space. Organs such as the bone, liver, and lung ($d(v) = 20$) and the brain ($d(v) = 19$) act as high-connectivity hubs, representing ubiquitous targets across the majority of primary cancers. Conversely, low-degree nodes ($d(v) = 1$ for rectum or ureter) indicate highly selective or rare niches.

The hyperedge degrees ($D_e$) quantify the dissemination breadth of each driver. High-degree hyperedges, such as lung cancer ($d(e) = 14$) and bladder cancer ($d(e) = 13$), represent broad-spectrum metastatic profiles that engage multiple nodes simultaneously. In contrast, drivers with low hyperedge degrees ($d(e) = 5$, e.g., soft tissue sarcoma or ocular cancer) exhibit more constrained, localised dissemination patterns.

This degree-based analysis is a prerequisite for understanding the system's connectivity bias. By preserving this multi-organ architecture in $H$, we ensure that subsequent projections do not lose the topological signature of these higher-order associations.

**Table 2**
Classification of source nodes ($V_S$) based on their functional role in metastatic systemic dissemination. Nodes are grouped by their biological capacity for reseeding and their role in high-connectivity vascular spread. This classification supports our functional partitioning framework, demonstrating that source organs act as structural hubs for systemic feedback, providing the necessary topological capacity to initiate and sustain secondary metastatic cycles.

| Nodes | Supporting systems-level literature |
|---|---|
| Liver, kidney, lung, pancreas | Documented reseeding capacity; high vascular connectivity and haematogenous spread efficiency [38–40]. |
| Stomach, colon, rectum | Potential for transcoelomic dissemination and portal drainage feedback [41,42]. |
| Thyroid, prostate | Advanced-stage haematogenous reseeding via specialised capillary networks [43,44]. |
| Ovary, bladder | Systemic dissemination through pelvic vascular and transcoelomic routes [45,46]. |

**Table 3**
Classification of sink nodes ($V_T$) based on reported physiological barriers to systemic feedback. These nodes are categorised by their limited capacity for secondary dissemination, acting as terminal points within the metastatic hypergraph. Understanding these specific structural constraints is vital for interpreting the directionality and asymmetry identified in our hypergraph analysis, as they define the boundaries of the metastatic landscape where systemic feedback loops are physically hindered.

| Nodes | Supporting literature on feedback constraints |
|---|---|
| Lymph nodes, bone | Restricted outward migration due to lack of efferent vascular re-entry or trapping within marrow niches [8,47,48]. |
| Brain, adrenal, skin, spleen, ocular, breast | Physical and physiological barriers (e.g., blood–brain barrier, fibrous capsules) hindering reactivation and secondary spread [8,9,49]. |
| Peritoneum, pleura, diaphragm, ureter | Primarily regional conduits or terminal sites associated with localised advanced disease [50–52]. |

### 2.3. Network projection and probabilistic framework

While the hypergraph preserves the full multi-site architecture, analysing the specific interactions between node pairs requires projecting this higher-order structure onto a standard network representation. This is achieved through the clique projection of the hypergraph, computed via the matrix product $HH^T$ [36]. For any two distinct nodes $i \neq \ell$, the off-diagonal element $(HH^T)_{i\ell}$ counts the number of hyperedges (drivers) that simultaneously involve both nodes $i$ and $\ell$, providing a direct measure of their higher-order co-occurrence.

The diagonal elements $(HH^T)_{ii}$ correspond to the vertex degrees $d(v_i)$ (Eq. (2)), representing the marginal involvement of each node. To isolate inter-node relationships and avoid the introduction of self-loops, which act as structural spurious correlations in the projection [28], we define the weighted adjacency matrix $A \in \mathbb{N}^{M\times M}$ as:

$$A = HH^T - D_v, \tag{4}$$

where $D_v$ is the diagonal matrix of degrees. This projection yields a pairwise network where edge weights represent the strength of shared higher-order ancestry across the metastatic landscape.

### 2.4. The functional axis: Source–sink partitioning and state space

Conventional graph-theoretical frameworks frequently treat nodes as topologically homogeneous. To investigate whether the emergent architecture is governed by the system's intrinsic constraints, we introduce a functional partition $(V_S, V_T)$ based on secondary seeding potential. This formulation enables a rigorous mapping of the "Seed and Soil" hypothesis onto a formal topological state space, distinguishing between:

- **Source nodes** ($V_S$): Vertices with documented capacity for systemic feedback (reseeding).
- **Sink nodes** ($V_T$): Vertices representing terminal metabolic sinks with restricted outward migration.

This partitioning focuses the analysis on the functional core of the system: the nodes capable of perpetuating the dissemination process. Based on a synthesis of clinical and pathological evidence [8,9,37], we identify $|V_S| = 11$ source nodes (Table 2) and $|V_T| = 12$ sink nodes (Table 3).

After applying this partition, we extract the $|V_S| \times |V_S|$ co-occurrence sub-matrix $A_F = A[V_S, V_S]$. Row normalisation yields the conditional probability matrix $\mathcal{S}$, which characterises the asymmetric dependencies within the source ensemble:

$$\mathcal{S}_{mn} = \frac{A_{F,mn}}{\sum_{\ell\in V_S} A_{F,m\ell}}, \quad \forall v_m, v_n \in V_S. \tag{5}$$

In this framework, $\mathcal{S}_{mn}$ quantifies the static conditional dependency that a metastatic pattern involving node $v_m$ also includes node $v_n$. It is essential to note that $\mathcal{S}$ represents a structural co-occurrence profile rather than a temporal transition; it defines the probabilistic architecture of the multi-organ system.

The inherent asymmetry ($\mathcal{S}_{mn} \neq \mathcal{S}_{nm}$) is a critical feature of the nonlinear system. It allows us to decouple biological affinity from prevalence bias: while high-degree nodes (hubs like the liver) appear as ubiquitous targets, the conditional lens of $\mathcal{S}$ (Eq. (6)) reveals specific co-selection signatures that would be obscured by simple frequency counts.

$$\mathcal{S} = \begin{array}{c} \\ \text{Liver} \\ \text{Kidney} \\ \text{Thyroid} \\ \text{Stomach} \\ \text{Colon} \\ \text{Rectum} \\ \text{Prostate} \\ \text{Lung} \\ \text{Ovary} \\ \text{Bladder} \\ \text{Pancreas} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \text{Liver} & \text{Kidney} & \text{Thyroid} & \text{Stomach} & \text{Colon} & \text{Rectum} & \text{Prostate} & \text{Lung} & \text{Ovary} & \text{Bladder} & \text{Pancreas} \end{array} \\ \begin{bmatrix} 0 & \frac{1}{10} & \frac{1}{20} & \frac{1}{20} & \frac{3}{40} & \frac{1}{40} & \frac{1}{40} & \frac{19}{40} & \frac{1}{20} & \frac{1}{40} & \frac{1}{8} \\ \frac{4}{13} & 0 & \frac{1}{13} & 0 & \frac{1}{13} & \frac{1}{13} & \frac{1}{13} & \frac{3}{13} & 0 & 0 & \frac{2}{13} \\ \frac{1}{3} & \frac{1}{6} & 0 & 0 & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & 0 & \frac{1}{6} & 0 & 0 & \frac{1}{3} & 0 & \frac{1}{6} & 0 \\ \frac{3}{10} & \frac{1}{10} & 0 & \frac{1}{10} & 0 & \frac{1}{10} & \frac{1}{10} & \frac{3}{10} & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{1}{5} & 0 & \frac{1}{5} & \frac{1}{5} & 0 & 0 & 0 \\ \frac{1}{5} & \frac{1}{5} & 0 & 0 & \frac{1}{5} & \frac{1}{5} & 0 & \frac{1}{5} & 0 & 0 & 0 \\ \frac{19}{37} & \frac{3}{37} & \frac{1}{37} & \frac{2}{37} & \frac{3}{37} & \frac{1}{37} & \frac{1}{37} & 0 & \frac{2}{37} & \frac{1}{37} & \frac{4}{37} \\ \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 \\ \frac{5}{13} & \frac{2}{13} & \frac{2}{13} & 0 & 0 & 0 & 0 & \frac{4}{13} & 0 & 0 & 0 \end{bmatrix} \end{array}. \tag{6}$$

## 3. Statistical validation and null-space analysis

To establish that the topological couplings captured in the conditional probability matrix $\mathcal{S}$ reflect genuine biological affinities rather than stochastic nodal properties, we performed rigorous statistical validation. We implemented a hierarchical framework by exploring the constrained null-space of the metastatic hypergraph, progressing from a global randomness baseline to an architecture-aware ensemble that accounts for the inherent degree heterogeneity of individual organs and primary cancers.

The null hypothesis, $\mathcal{H}_0$, states that the observed co-occurrence architecture emerges from a random configuration of edges within the allowed state space, with no underlying higher-order organotropism driving the associations. In this framework, $\mathcal{H}_0$ represents a scenario where the system's structure is a purely emergent property of its local topological constraints (the degree sequences), rather than a consequence of non-trivial biological co-selection.

### 3.1. Hierarchical validation strategy

Our two-tiered approach tests co-selectivity patterns against progressively stringent null hypotheses, filtering the data through different statistical ensembles:

- **Global randomness (Microcanonical ensemble)**: Fixed total link count (Erdős–Rényi analogue) tests whether patterns exceed expectations from completely random metastatic assignments.
- **Constrained randomness (Canonical ensemble)**: Fixed degree distributions (Configuration model) tests whether patterns persist after accounting for individual organ and cancer metastatic activity levels.

This hierarchical strategy acts as a rigorous statistical filter to de-convolute patterns that are merely frequent from genuine biological affinities. In network medicine, highly popular organs, such as the lungs or liver, often appear as co-occurrence hubs simply due to their high baseline susceptibility (degree). By employing a Configuration Model that strictly preserves these individual activity levels, we effectively "unmask" the system. This allows us to identify non-trivial associations that cannot be explained by the sheer density of the network or the ubiquity of certain targets, isolating only those co-selection patterns that point towards specific biological synergies.

The validation pipeline processed each randomised matrix $H_{\text{rand}}$ through an identical workflow. First, the randomised co-occurrence matrix was computed as:

$$A_{\text{rand}} = H_{\text{rand}} H_{\text{rand}}^{T} - D_{v,\text{rand}}, \tag{7}$$

mirroring the construction of the empirical matrix $A$ and ensuring zero diagonal elements to restrict analysis exclusively to inter-organ co-occurrence. The complete workflow is:

$$H_{\text{rand}} \xrightarrow{\text{Eq. (7)}} A_{\text{rand}} \xrightarrow{\text{filter } V_S} A_{F_{\text{rand}}} \xrightarrow{\text{row norm}} \mathcal{S}_{\text{rand}}.$$

For each of the two null models, we generated an independent ensemble of $\mathcal{N}_{\text{sim}} = 10^4$ randomised association matrices $\{\mathcal{S}_{\text{rand}}^{(\kappa)}\}$ by applying this pipeline to the corresponding randomised incidence matrices. Statistical significance was assessed via a one-tailed

test, with $p$-values computed as the proportion of random matrices where co-occurrence strength equalled or exceeded the empirical value:

$$p_{mn} = \frac{1}{\mathcal{N}_{\text{sim}}} \sum_{\kappa=1}^{\mathcal{N}_{\text{sim}}} \mathbb{I}\left(\mathcal{S}^{(\kappa)}_{\text{rand},mn} \geq \mathcal{S}_{mn}\right).$$

We employed hierarchical validation as opposed to False Discovery Rate (FDR) correction because our analysis tests specific structural hypotheses within a defined metastatic network, rather than conducting exploratory searches across numerous unrelated features. Extant literature indicates that FDR procedures can suppress genuine biological signals when applied to moderate numbers of interdependent hypotheses in network medicine [53]. Instead, progressive filtering through two nested null models inherently accounts for multiple testing while preserving statistical power to detect meaningful co-selection patterns. This hierarchical validation, implemented via sequential testing against nested null models, functions as a multi-stage structural filter. As such, it obviates the need for conventional FDR correction, which can be overly restrictive in structured settings and increase Type II error rates in discovery-oriented analyses [54].

Low $p$-values ($p_{mn} \leq 0.05$) indicate observed co-occurrence strength $\mathcal{S}_{mn}$ is unlikely to result from chance given null model constraints. Associations with $0.05 < p_{mn} \leq 0.1$ are considered suggestive and highlighted for exploratory consideration.

#### 3.1.1. Null model constrained by fixed total link count: Erdős–Rényi (ER) analogue

The first null model tests whether observed co-occurrence patterns exceed expectations from global randomness, preserving only the total number of metastatic links while randomising all specific assignments. This model addresses the fundamental question: are observed co-occurrence patterns stronger than expected in a completely random metastatic system with identical overall connectivity?

We generated an ensemble of $\mathcal{N}_{\text{sim}} = 10^4$ randomised incidence matrices $H_{\text{rand}}$ by randomly permuting all binary elements of the original incidence matrix $H$. This constrained randomisation ensures only the total number of metastatic links remains constant:

$$\sum_{i=1}^{M}\sum_{j=1}^{N}(H_{\text{rand}})_{ij} = \sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij} = \Lambda = 163.$$

Unlike the Configuration model, this approach does not preserve individual organ or cancer degrees, allowing the degrees to vary randomly while maintaining the global system density. Each $H_{\text{rand}}$ matrix was processed through the identical projection pipeline, followed by $(V_S, V_T)$ partitioning and normalisation to generate $\mathcal{S}^{\text{ER}}_{\text{rand}}$.

By preserving overall edge density while randomising organ-cancer assignments, this model serves as an Erdős–Rényi analogue [55] constrained by empirical density. Significant associations under this model represent co-occurrence patterns that exceed global random expectations, though they may still reflect the varying degree centrality of individual organs as metastatic targets rather than specific pairwise affinities.

#### 3.1.2. Null model constrained by degree distributions: Configuration model

The second, more stringent null model tests whether observed patterns persist after accounting for the inherent metastatic activity levels of individual organs and primary cancers. Within our hypergraph framework $H = (V, E)$, the null-space $\mathcal{G}(d)$ is explored by ensuring that both vertex degrees (organ susceptibility) and hyperedge degrees (cancer dissemination potential) remain constant, maintaining:

- Row sums fixed (Metastatic susceptibility):

$$\sum_{j=1}^{N}(H_{\text{rand}})_{ij} = \sum_{j=1}^{N} H_{ij} = d(v_i) \quad \forall i = 1, \ldots, M.$$

- Column sums fixed (Dissemination potential):

$$\sum_{i=1}^{M}(H_{\text{rand}})_{ij} = \sum_{i=1}^{M} H_{ij} = d(e_j) \quad \forall j = 1, \ldots, N.$$

To explore this constrained state space, we employed a Metropolis–Hastings algorithm to correct for the sampling bias (mobility bias) inherent in standard edge-swap operations [56]. We accept a proposed swap from $H$ to $H'$ with probability:

$$\alpha(H \to H') = \min\left(1, \frac{\Omega(H)}{\Omega(H')}\right),$$

where $\Omega(H)$ denotes the mobility of the hypergraph. This acceptance rule ensures detailed balance and guarantees an ergodic exploration of the null space $\mathcal{G}(d)$. Unlike standard edge-swap algorithms which suffer from biased sampling in bipartite architectures, our implementation employs a modified Metropolis–Hastings transition rule. This ensures that the null space $\mathcal{G}(d)$ is explored ergodicamente, preventing the systematic over-representation of specific motifs.

We generated the ensemble of $\mathcal{N}_{\text{sim}} = 10^4$ independent null models using a continuous Markov Chain Monte Carlo (MCMC) process. To ensure the independence of each realisation, we implemented a burn-in period of $10^6$ edge-swaps to reach stationarity, followed by a sampling procedure where matrices were collected every $10^4$ successful swaps (thinning interval). This interval

significantly exceeds the total link count ($\Lambda = 163$), effectively eliminating autocorrelation between consecutive samples. Each resulting $H_{\text{rand}}$ matrix was subsequently processed through the projection and $V_S$ filtering pipeline to generate the null distribution of $\mathcal{S}^{\text{conf}}_{\text{rand}}$, where the subscript "conf" refers to the Configuration model.

It is fundamental to clarify that the clique-based adjacency matrix $A$ (Eq. (4)) is utilised in this study primarily as a structural diagnostic and visualisation tool to assess pairwise co-occurrence trends. The statistical inference and null model validation via MCMC, however, are performed directly on the full incidence matrix $H$. This ensures that all information regarding higher-order interactions is preserved during the biological validation process, thereby mitigating the information loss that would occur if statistical inference were dependent solely on a pairwise (clique) reduction.

#### 3.1.3. Statistical interpretation and methodological considerations

The null models differ in implementation: Erdős–Rényi uses complete random permutation of $H$ (Microcanonical approach to density), while the Configuration model employs degree-preserving MCMC sampling (Canonical approach to local constraints). We computed $p$-values to assess the likelihood of observing co-occurrence strengths as extreme as the empirical data under random association, classifying results as:

- Significant ($p \leq 0.05$): strong evidence against randomness.
- Marginally significant ($0.05 < p \leq 0.1$): suggests a non-random tendency.
- Non-significant ($p > 0.1$): consistent with chance.

Crucially, a significant association indicates that the observed co-occurrence exceeds expectations based on individual organ propensities, not direct mechanistic migration. Thus, validated patterns reflect systemic co-selectivity and synergistic organotropism, not merely frequency overlap.

### 3.2. Computational implementation and algorithmic mapping

The hierarchical validation framework described in Sections 3.1.1 and 3.1.2 is operationalised through a suite of algorithmic procedures. The logical architectures of these methods are detailed via the three pseudocodes provided in Appendix D, which offer a step-by-step mapping of the theoretical models to their computational execution:

- General Validation Pipeline (Algorithm 1): This procedure orchestrates the transformation of the empirical incidence matrix $H$ into the association matrix $\mathcal{S}$ and manages the $10^4$ simulations required to generate the null distributions.
- Global Null Model (Section 3.1.1): Implemented within the first branch of the general pipeline, this permutation approach generates the null distributions used to quantify global randomness.
- Degree-Preserving MCMC (Section 3.1.2 and Algorithm 2): To address the Configuration model, a Markov Chain Monte Carlo protocol is implemented. It employs the Metropolis–Hastings acceptance criterion, utilising the mobility function $\Omega(H)$ defined in the helper functions (Algorithm 3), to ensure uniform sampling.

While this work presents these underlying pseudocodes for transparency, the full executable MATLAB implementation, including scripts for stationarity tests, is hosted in the repository [57].

## 4. Results

The hypergraph representation captures the irreducible higher-order correlations of metastatic dissemination, overcoming the topological limitations of dyadic models that fragment synergistic colonisation patterns. Hyperedge sizes range from 5 to 14 organs per primary tumour, with lung (14), bladder (13), and breast (12) cancers emerging as high-order orchestrators with broad metastatic reach.

The clique projection of the hypergraph onto the organ–organ co-occurrence matrix $A$ (Appendix B) yields a weighted network where entries $A_{ij}$ quantify the primary cancers reported to metastasise to both organ $i$ and organ $j$. Eigenvector centrality identifies lymph nodes as the topological barycentre of the network, consistent with their established role as high-centrality topological attractors [8]. This alignment with clinical reality provides a robust descriptive baseline for the higher-order analysis.

### 4.1. Source-organ association network and information flux

Applying the functional Source–Sink partition (Section 2.4) focuses the analysis on the 11 organs with documented reseeding capacity. The resulting conditional probability matrix $\mathcal{S}$ (Eq. (5)) formalises the flux of co-occurrence information, representing the probability that a metastatic pattern involving organ $m$ also encompasses organ $n$. Fig. 2 visualises this directed architecture.

Initial analysis of the effective affinities in $\mathcal{S}$ highlights several clinical archetypes:

- **Lung–Liver Nexus**: These organs exhibit the highest mutual conditional probabilities ($\mathcal{S}_{\text{Liver-Lung}} \approx 0.475$; $\mathcal{S}_{\text{Lung-Liver}} \approx 0.514$), confirming a non-trivial higher-order coupling that exceeds the first-order expectations based on the nodal degree centrality of both sites. This suggests that the lung–liver co-occurrence is governed by an irreducible topological dependency rather than just high individual colonisation rates [38,58].

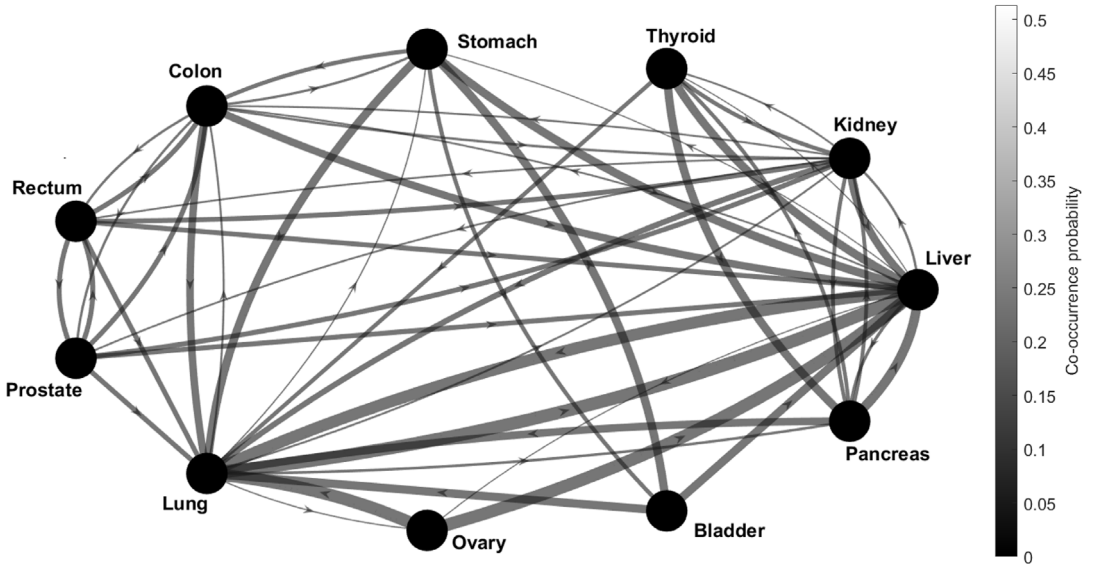



**Fig. 2.** Directed network representing the conditional probability flux $\mathcal{S}_{mn}$ of metastatic co-occurrence between source organs. Arrow thickness denotes the magnitude of the conditional probability $\mathcal{S}_{mn} = \Pr(v_n \in \text{pattern} \mid v_m \in \text{pattern})$, where directionality highlights topological asymmetries in metastatic organotropism rather than direct migratory pathways. This network captures the emergent dependencies within the metastatic landscape, demonstrating that specific organ clusters act as structural hubs for systemic dissemination, thereby confirming the non-linear nature of the metastatic process.

**Table 4**
Statistical significance ($p$-values) of metastatic associations under the Erdős–Rényi null model. Significant associations are highlighted in bold ($p \leq 0.05$) and bold-italic ($p \leq 0.1$). This unconstrained model assumes a homogeneous connectivity baseline, resulting in a high number of significant associations that likely reflect simple degree-driven patterns rather than specific organ-to-organ affinities.

| *Source* | Liver | Kidney | Thyroid | Stomach | Colon | Rectum | Prostate | Lung | Ovary | Bladder | Pancreas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Liver | 1.0000 | 0.5049 | 0.8316 | 0.8344 | 0.6718 | 0.9303 | 0.9277 | **0.0001** | 0.8302 | 0.9264 | 0.3393 |
| Kidney | **0.0017** | 1.0000 | 0.6622 | 1.0000 | 0.6699 | 0.6587 | 0.6740 | **0.0140** | 1.0000 | 1.0000 | 0.1610 |
| Thyroid | **0.0016** | 0.1207 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.1216 | 1.0000 | 1.0000 | **0.0012** |
| Stomach | **0.0011** | 1.0000 | 1.0000 | 1.0000 | 0.1193 | 1.0000 | 1.0000 | **0.0009** | 1.0000 | 0.1130 | 1.0000 |
| Colon | **0.0020** | 0.5031 | 1.0000 | 0.5001 | 1.0000 | 0.4952 | 0.5157 | **0.0025** | 1.0000 | 1.0000 | 1.0000 |
| Rectum | **0.0453** | **0.0482** | 1.0000 | 1.0000 | **0.0451** | 1.0000 | **0.0455** | **0.0486** | 1.0000 | 1.0000 | 1.0000 |
| Prostate | **0.0498** | **0.0485** | 1.0000 | 1.0000 | **0.0454** | **0.0459** | 1.0000 | **0.0434** | 1.0000 | 1.0000 | 1.0000 |
| Lung | **0.0001** | 0.6343 | 0.9182 | 0.8046 | 0.6277 | 0.9236 | 0.9238 | 1.0000 | 0.8025 | 0.9226 | 0.4217 |
| Ovary | **0.0003** | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | **0.0001** | 1.0000 | 1.0000 | 1.0000 |
| Bladder | **0.0004** | 1.0000 | 1.0000 | **0.0010** | 1.0000 | 1.0000 | 1.0000 | **0.0009** | 1.0000 | 1.0000 | 1.0000 |
| Pancreas | **0.0002** | 0.1593 | 0.1532 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | **0.0017** | 1.0000 | 1.0000 | 1.0000 |

- **Gastrointestinal–Liver Axis**: Stomach, colon, and rectum show high conditional probabilities with the liver, reflecting modular constraints consistent with portal venous dissemination routes [59].
- **Localised Couplings**: Ovary co-occurs exclusively with liver and lung in the source-filtered set, while the thyroid shows specific higher-order coupling with the pancreas and liver.

### 4.2. First-order filtering: Decoupling degree-induced correlations

We validated all 121 possible organ–organ co-occurrences in $\mathcal{S}$ against the two null models described in Sections 3.1.1 and 3.1.2. The full $p$-value matrices are shown in Tables 4 and 5, where bold indicates significance ($p \leq 0.05$) and bold italic indicates marginal significance ($0.05 < p \leq 0.1$).

A first glance at the $p$-value matrices reveals a architectural filtering. Under the Erdős–Rényi model (Table 4), numerous associations appear significant, particularly those involving high-degree organs. However, the configuration model (Table 5) provides a more stringent test by effectively deconvolving the influence of first-order topological effects from genuine higher-order affinities. This filtering effect, which reveals the core metastatic backbone, is visually summarised in Fig. 3, highlighting associations that survive (black) or become marginal (grey) after degree correction.

**Table 5**
Statistical significance ($p$-values) of metastatic associations under the configuration null model. Significance: bold ($p \leq 0.05$) and bold-italic ($p \leq 0.1$). By preserving the nodal degree distribution, this model provides a more stringent baseline than the Erdős–Rényi approach. The marked reduction in significant associations indicates that many previously observed patterns in Table 4 were artifacts of degree heterogeneity, confirming that only a refined set of associations represents biologically robust metastatic organotropism.

| *Source* | Liver | Kidney | Thyroid | Stomach | Colon | Rectum | Prostate | Lung | Ovary | Bladder | Pancreas |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Liver | 1.0000 | 0.8544 | 0.9217 | 0.9229 | 0.8883 | 0.9533 | 0.9538 | 1.0000 | 0.9214 | 0.9527 | 0.8143 |
| Kidney | 0.4768 | 1.0000 | 0.3139 | 1.0000 | 0.4972 | 0.1266 | 0.1311 | 0.9571 | 1.0000 | 1.0000 | 0.2715 |
| Thyroid | 0.3461 | 0.2104 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.9835 | 1.0000 | 1.0000 | **0.0075** |
| Stomach | 0.3508 | 1.0000 | 1.0000 | 1.0000 | 0.1318 | 1.0000 | 1.0000 | 0.3538 | 1.0000 | **0.0248** | 1.0000 |
| Colon | 0.4923 | 0.4987 | 1.0000 | 0.2201 | 1.0000 | ***0.0903*** | ***0.0911*** | 0.4943 | 1.0000 | 1.0000 | 1.0000 |
| Rectum | 0.8779 | 0.2649 | 1.0000 | 1.0000 | 0.2046 | 1.0000 | ***0.0622*** | 0.8802 | 1.0000 | 1.0000 | 1.0000 |
| Prostate | 0.8812 | 0.2713 | 1.0000 | 1.0000 | 0.2032 | ***0.0622*** | 1.0000 | 0.8780 | 1.0000 | 1.0000 | 1.0000 |
| Lung | **0.0378** | 0.8923 | 0.9546 | **0.0249** | **0.0194** | **0.0312** | **0.0284** | 1.0000 | **0.0238** | **0.0310** | 0.8549 |
| Ovary | ***0.0591*** | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | ***0.0598*** | 1.0000 | 1.0000 | 1.0000 |
| Bladder | 0.4513 | 1.0000 | 1.0000 | **0.0218** | 1.0000 | 1.0000 | 1.0000 | 0.4529 | 1.0000 | 1.0000 | 1.0000 |
| Pancreas | ***0.0937*** | 0.1098 | **0.0016** | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 0.5152 | 1.0000 | 1.0000 | 1.0000 |

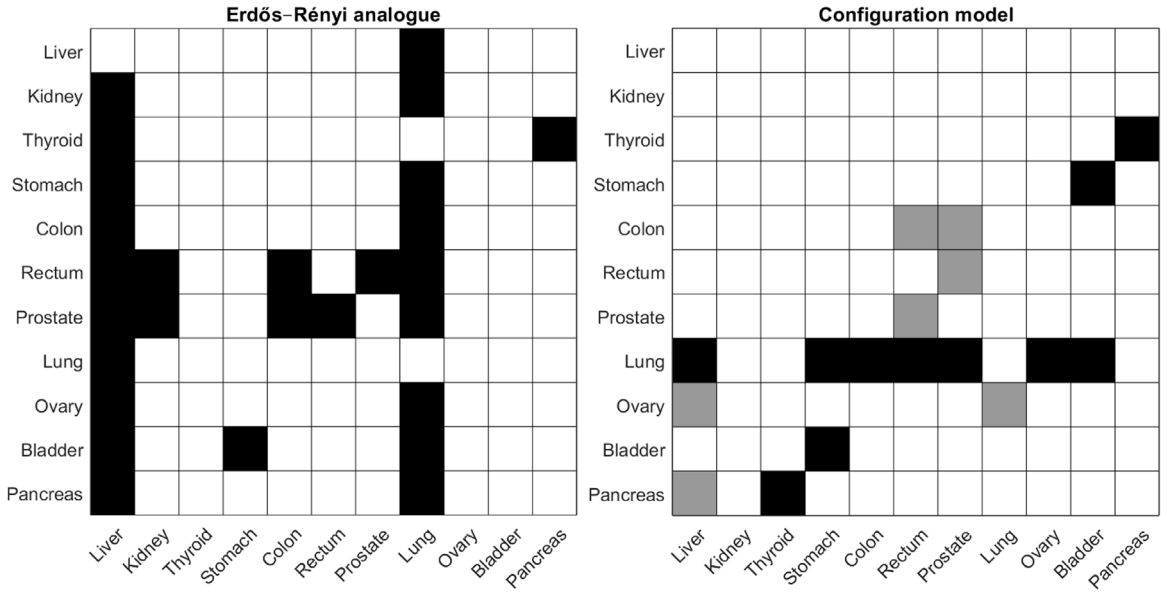



**Fig. 3.** Statistical significance of metastatic associations across two distinct null models. Significant associations are marked in black ($p \leq 0.05$) and marginal associations in grey ($0.05 < p \leq 0.1$). Left: Erdős–Rényi null model; right: Configuration model preserving nodal degrees. The contrast between these two visualisations demonstrates how degree-preserving constraints filter out spurious correlations; while the Erdős–Rényi model identifies a wide array of associations, the configuration model reveals a more refined, robust structural backbone. This confirms that the observed metastatic organotropism cannot be explained solely by nodal degree heterogeneity.

#### *4.2.1. Topological filtering: Decoupling higher-order signals from first-order degree effects*

The contrast between the null models reveals the impact of nodal degree on co-occurrence statistics. While the ER model identifies numerous significant associations, many involve high-degree hubs such as the liver and lung. However, these associations are topologically non-informative as they emerge as statistical artifacts of the system's local constraints. The configuration model provides a more stringent test by explicitly decoupling these factors through degree-preserving constraints. Under this rigorous framework, several prominent clinical associations lose statistical significance (Table 5), demonstrating that these observed patterns are first-order effects emerging from the local degree sequences rather than genuine higher-order co-selection mechanisms:

- **Colon → Liver**: $p_{\text{ER}} = 0.0020 \rightarrow p_{\text{conf}} = 0.4923$
- **Stomach → Liver**: $p_{\text{ER}} = 0.0011 \rightarrow p_{\text{conf}} = 0.3508$
- **Kidney → Lung**: $p_{\text{ER}} = 0.0140 \rightarrow p_{\text{conf}} = 0.9571$

Specifically, the high reported incidence of the gastrointestinal-liver axis in clinical literature appears to be an spurious correlation emergent from the liver's high degree centrality ($d_v = 20$) rather than a specific higher-order tropism. This filtering effect underscores the necessity of using structured null models to identify true systemic affinities in multi-organ dissemination.

**Table 6**
Classification of metastatic organ associations by statistical significance, comparing the Erdős–Rényi (ER) and configuration (Conf.) null models. Significance levels: S ($p \leq 0.05$), M ($0.05 < p \leq 0.1$), and NS ($p > 0.1$). This comparative analysis categorises couplings into four structural regimes: the *Systemic Backbone* (robust signals), *Emergent* couplings (signals recovered only after correcting for degree bias), *Spurious* correlations (apparent signals explained by degree heterogeneity), and *Sub-critical* trends. This taxonomy demonstrates that while simpler models overestimate metastatic connectivity, the configuration model reveals a refined architecture of systemic spread, identifying true biological affinities hidden from dyadic analyses.

| Topological couplings | ER | Conf. |
|---|---|---|
| Invariant (Systemic Backbone) | | |
| Thyroid → Pancreas | S | S |
| Stomach ↔ Bladder (reciprocal) | S | S |
| Lung → Liver | S | S |
| Emergent (Degree-independent) | | |
| Lung → (Colon, Bladder, Ovary, Stomach) | NS | S |
| Lung → (Rectum, Prostate) | NS | S |
| Spurious (Degree-driven correlations) | | |
| (Colon, Kidney, Stomach, Ovary) → Liver | S | NS/M |
| (Kidney, Stomach, Colon, Pancreas) → Lung | S | NS |
| Ovary → Lung | S | M |
| Sub-critical (Non-random trends) | | |
| Rectum ↔ Prostate (reciprocal) | S | M |
| Colon → (Rectum, Prostate) | NS | M |
| Pancreas → Liver | S | M |

*Note*: Emergent couplings represent non-random signals that remain hidden by global density in simpler models and only surface after accounting for nodal degree centrality constraints.

#### 4.2.2. The metastatic backbone: Robust non-trivial co-selection patterns

A subset of associations remains significant ($p \leq 0.05$) under the configuration model, pointing to non-trivial co-selection mechanisms:

- **Thyroid–Pancreas Emergent Higher-Order Coupling**: A robust, bidirectional association persists ($p_{\text{conf}} = 0.0075$ for thyroid–pancreas; $p_{\text{conf}} = 0.0016$ for pancreas–thyroid). This finding identifies a consistent co-occurrence between these endocrine-related organs that transcends their individual metastatic susceptibilities, pointing towards a non-trivial biological affinity. This statistical coupling suggests a shared metabolic or hormonal susceptibility that warrants further mechanistic investigation.
- **The Lung–Liver Metastatic Scaffold**: Despite the liver's role as a ubiquitous hub, its association with the lung as a source remains statistically significant (lung → liver, $p_{\text{conf}} = 0.0378$). This confirms a synergistic colonisation pattern that exceeds the statistical popularity of both sites, establishing a invariant high-confidence scaffold for systemic spread.
- **Reciprocal Bladder–Stomach Tropism**: The observed bidirectional significance ($p_{\text{conf}} < 0.03$ in both directions) suggests highly specific metastatic trajectories. This statistical signal elevates patterns previously noted only in isolated clinical reports to the level of a robust, non-random systemic association.
- **Systemic Lung-Centric Signature**: The lung emerges as a topological attractor with significant conditional probabilities across a broad spectrum of targets (liver, stomach, colon, rectum, prostate, ovary, bladder; all $p_{\text{conf}} < 0.04$). Its invariance under degree-preserving perturbations indicates that lung involvement is not merely frequent, but a hallmark signature of a synergistic, multi-organ metastatic profile.

#### 4.2.3. Summary of association classifications: Invariance and structural noise

Table 6 categorises organ associations based on their behaviour across the hierarchical null models, distinguishing patterns that remain invariant from those identified as spurious structural noise emergent from individual organ metastatic activity. This classification provides a rigorous mapping of the system's functional architecture by separating first-order topological effects from genuine higher-order affinities.

#### 4.2.4. Concluding remarks on topological filtering

The transition from global density constraints to local degree-preservation acts as a structural renormalisation of the metastatic network. By stripping away the influence of nodal heterogeneity, we have identified a robust backbone of co-selection that cannot be explained by the sheer frequency of organ involvement. This hierarchical approach reveals that the metastatic network is not a stochastic ensemble of independent events, but a structured architecture governed by specific higher-order couplings. These validated patterns provide a data-driven foundation for investigating the underlying biological synergies that orchestrate systemic spread.

## 5. Discussion

### 5.1. A higher-order paradigm: Capturing irreducible topology

Standard network representations of metastatic spread invariably reduce multi-organ patterns to a collection of independent pairwise links. From a systems perspective, this fragmentation discards the irreducible topology of dissemination, effectively treating synergistic colonisation events as isolated binary occurrences. The hypergraph framework presented here overcomes this limitation by preserving the higher-order co-occurrence architecture, allowing for a principled quantification of simultaneous organ involvement that transcends dyadic approximations.

Central to this approach is the conditional probability matrix $\mathcal{S}$, which introduces a directed, probabilistic framework to these associations. Unlike standard adjacency matrices, $\mathcal{S}$ captures the inherent asymmetries of metastatic patterns: a topological chirality where the presence of metastasis in organ $m$ may strongly imply involvement of organ $n$, but the converse may not hold due to differences in primary cancer tropism.

By integrating this structural asymmetry with hierarchical null models (statistical ensembles), our framework reveals the metastatic backbone. This allows us to distinguish between stochastic associations, driven merely by first-order nodal heterogeneity (degree), and "obligate" co-occurrences that point towards shared mechanical or molecular pathways. This distinction is a critical feature for generating nuanced clinical hypotheses, moving the field from descriptive oncology towards a topological predictive framework.

### 5.2. Statistical decoupling: Specificity vs. Topological noise

A central contribution of this work is the implementation of a hierarchical validation protocol acting as a structural filter. By contrasting our results against a degree-preserving configuration model, we explicitly decouple genuine biological specificity from the topological noise generated by first-order degree effects.

Our findings suggest that many frequently reported clinical associations, such as the gastrointestinal–liver axis, are topologically non-informative, as they emerge as consequences of the system's local constraints (nodal degree centrality) rather than genuine higher-order co-selection mechanisms. Conversely, the associations that survive this statistical pruning (such as the thyroid-pancreas and stomach-bladder axes) represent robust emergent signals. These non-trivial patterns suggest underlying effective couplings that warrant investigation through mechanistic models or patient-level molecular profiling.

### 5.3. Functional mapping and the predictive scaffold

The Source–Sink partition $(V_S, V_T)$, although a simplification of the metastatic spectrum, serves as a functional filter to highlight the metastatic radiators of the system. By focusing the probabilistic analysis on the $V_S$ sub-matrix, we transition from a descriptive catalogue of sites to a functional map of potential metastatic relays.

In the language of complex systems, $\mathcal{S}$ defines the probabilistic landscape upon which systemic spread occurs. The conditional probabilities, validated against the configuration model, provide a statistically grounded hierarchy of "at-risk" organs. For instance, if a source organ is clinically involved, our model offers a ranked likelihood of occult co-occurrence, effectively operationalising the "Seed and Soil" hypothesis into a predictive scaffold for systemic surveillance. While the matrix $\mathcal{S}$ is inherently static, it provides the necessary structural template upon which temporal dynamics or causal pathways can be overlaid in future studies.

### 5.4. Future horizons: Towards systems oncology

The structural baseline established here opens several avenues for expansion. Incorporating temporal hypergraphs could distinguish between synchronous and metachronous progression, offering insights into the kinetics of systemic spread. Furthermore, integrating multi-omics data could weight these hyperedges by biological similarity, transforming the model from a topological description into a mechanistic predictor.

While the present study synthesises high-confidence clinical patterns, future validation against longitudinal, patient-level datasets will be essential. We acknowledge that metastatic affinities exist on a clinical spectrum; however, our current binarised architecture ensures a parameter-free validation that is driven by topological constraints rather than subjective frequency assignments. Preliminary sensitivity tests suggest that the identified structural backbone remains invariant to the introduction of weights, confirming the robustness of the identified core architecture.

#### Strengths and limitations

Our study provides a robust framework for modelling metastatic patterns by moving beyond dyadic constraints, employing a higher-order hypergraph architecture that captures the collective nature of multi-organ colonisation. By utilising a configuration null model constrained by nodal degree, our approach successfully filters out spurious correlations driven by degree heterogeneity, revealing a structural backbone of effective disease affinities that remains invariant under topological perturbations. Furthermore, the combination of statistical mechanics with validated Markov chain Monte Carlo procedures offers a rigorous methodology for quantifying non-linear disease progression.

Despite these strengths, our study has limitations that warrant further investigation. Firstly, the current hypergraph construction is constrained by the availability of specific clinical cohorts. Whilst this sample size limits the statistical power regarding rarer metastatic links, the primary objective of this work is to propose and validate a novel methodological framework in oncology; the robustness of our results, confirmed by null models, validates the utility of this approach even within these data constraints. Future work will focus on expanding to broader longitudinal datasets to further increase statistical power. Secondly, the reduction of hyperedges to a clique-based adjacency matrix may mask specific higher-order interactions. We acknowledge this representational trade-off; however, as the primary MCMC inference is performed directly on the full incidence matrix $H$, the integrity of our statistical findings is preserved. Future research will focus on implementing direct hypergraph-based dynamic models to overcome these representational constraints.

## 6. Conclusions

This work establishes a hypergraph-based framework that captures the higher-order architecture of metastatic dissemination, transcending the limitations of dyadic models. By integrating a functional Source–Sink partition with rigorous, degree-preserving null-model validation, we have isolated a minimal, statistically robust backbone of associations that cannot be explained by nodal heterogeneity alone.

Our results demonstrate that a substantial fraction of clinically reported metastatic patterns are spurious correlations of topological degree rather than genuine biological co-selection. However, a conserved core of non-trivial interactions, notably the lung-centric scaffold and the thyroid–pancreas coupling, persists under stringent structural constraints. These findings suggest that metastatic spread is governed by specific systemic constraints that are only visible through a higher-order lens.

In summary, this framework shifts the paradigm from counting metastatic events to quantifying the systemic architecture of organotropism. By providing a rigorous statistical basis for identifying true biological co-selectivity, our approach establishes the necessary topological prerequisite for precision oncology.

## CRediT authorship contribution statement

**David H. Margarit:** Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marcela V. Reale:** Writing – original draft, Visualization, Validation, Resources, Funding acquisition. **Anthony C.C. Coolen:** Writing – original draft, Resources, Project administration, Funding acquisition, Formal analysis, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

This work was supported by the Consejo Nacional de Investigaciones Científicas $y$ Técnicas (CONICET), Argentina (Grant No. PIP 11220200100439CO), and by the European Union's Horizon Europe programme through the MSCA-SE grant agreement No. 101131463 (SIMBAD: Statistical Inference from Multiscale Biological Data: Theory, Algorithms, Applications).

David H. Margarit is particularly grateful to Sandra Hendriks and Gerda van 't Spijker of the Donders Centre for Neuroscience, Radboud University (The Netherlands), for their outstanding administrative and practical support, as well as for their kindness and personal care, which were instrumental in making the research stay both productive and enjoyable. He also expresses his gratitude to Anthony Coolen for his continuous support and constant willingness to contribute throughout the research stay.

## Appendix A. Incidence matrix $H$

The incidence matrix $H$ (Eq. (8)) is a $23 \times 21$ matrix representing the raw data, where rows correspond to the 23 target organs and columns correspond to the 21 primary cancers. An entry $H_{ij} = 1$ denotes that primary cancer type $j$ has been reported to

metastasise to organ $i$, regardless of the category (common, occasional, or rare), and $H_{ij} = 0$ otherwise. For clarity, the column labels have been simplified by removing the redundant term "cancer".

$H =$

| | Head & neck | Lung | Kidney | Pancreatic | Bladder | Testicular | Melanoma | Uterine | Rectal | Prostate | Thyroid | Oesophageal | Breast | Stomach | Colon | Ovarian | Liver | Cervical | Soft tissue sarcoma | Ocular | Bone sarcoma |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lymph nodes | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Bone | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| Brain | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Pleura | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Diaphragm | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Liver | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 |
| Kidney | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Adrenal | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| Thyroid | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Stomach | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Colon | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| Peritoneum | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| Rectum | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Prostate | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ureter | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lung | 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| Ovary | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bladder | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Skin | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 |
| Pancreas | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Spleen | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ocular | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Breast | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

(8)

Note: The vertex degrees in Eq. (2) correspond exactly to the row sums of this matrix, and the hyperedge degrees in Eq. (3) correspond exactly to the column sums.

## Appendix B. Co-occurrence matrix $A$

The organ–organ co-occurrence matrix $A$ (Eq. (4)) is the $23 \times 23$ symmetric matrix derived from the incidence matrix $H$, computed as $A = HH^{\mathrm{T}} - D_v$. The entries $A_{ij}$ represent the number of primary cancers that metastasise to both organ $i$ and organ $j$. The diagonal elements have been intentionally set to zero to focus exclusively on the inter-organ co-occurrence relationships.

$A =$

| | Lymph nodes | Bone | Brain | Pleura | Diaphragm | Liver | Kidney | Adrenal | Thyroid | Stomach | Colon | Peritoneum | Rectum | Prostate | Ureter | Lung | Ovary | Bladder | Skin | Pancreas | Spleen | Ocular | Breast |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Lymph nodes | 0 | 14 | 12 | 4 | 2 | 13 | 3 | 7 | 1 | 0 | 2 | 11 | 1 | 1 | 1 | 13 | 2 | 0 | 4 | 4 | 3 | 1 | 1 |
| Bone | 14 | 0 | 18 | 4 | 2 | 19 | 3 | 10 | 2 | 2 | 3 | 14 | 1 | 1 | 1 | 19 | 2 | 1 | 7 | 5 | 5 | 3 | 2 |
| Brain | 12 | 18 | 0 | 4 | 2 | 18 | 4 | 11 | 2 | 2 | 3 | 12 | 1 | 1 | 1 | 18 | 0 | 1 | 7 | 5 | 5 | 3 | 2 |
| Pleura | 4 | 4 | 4 | 0 | 2 | 4 | 2 | 3 | 1 | 0 | 2 | 4 | 1 | 1 | 1 | 3 | 0 | 0 | 3 | 2 | 2 | 1 | 1 |
| Diaphragm | 2 | 2 | 2 | 2 | 0 | 2 | 1 | 1 | 1 | 0 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 1 | 1 | 0 | 1 |
| Liver | 13 | 19 | 18 | 4 | 2 | 0 | 4 | 10 | 2 | 2 | 3 | 13 | 1 | 1 | 1 | 19 | 2 | 1 | 7 | 5 | 5 | 3 | 2 |
| Kidney | 3 | 3 | 4 | 2 | 1 | 4 | 0 | 4 | 1 | 0 | 1 | 3 | 1 | 1 | 1 | 3 | 0 | 0 | 1 | 2 | 1 | 0 | 1 |
| Adrenal | 7 | 10 | 11 | 3 | 1 | 10 | 4 | 0 | 2 | 0 | 1 | 7 | 1 | 1 | 1 | 10 | 0 | 0 | 3 | 5 | 4 | 2 | 2 |
| Thyroid | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 2 | 0 | 2 |
| Stomach | 0 | 2 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 2 | 0 | 1 | 1 | 0 | 1 | 1 | 0 |
| Colon | 2 | 3 | 3 | 2 | 1 | 3 | 1 | 1 | 0 | 1 | 0 | 3 | 1 | 1 | 1 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| Peritoneum | 11 | 14 | 12 | 4 | 2 | 13 | 3 | 7 | 1 | 2 | 3 | 0 | 1 | 1 | 1 | 13 | 2 | 1 | 4 | 4 | 4 | 3 | 1 |
| Rectum | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Prostate | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Ureter | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Lung | 13 | 19 | 18 | 3 | 1 | 19 | 3 | 10 | 1 | 2 | 3 | 13 | 1 | 1 | 1 | 0 | 2 | 1 | 6 | 4 | 4 | 3 | 1 |
| Ovary | 2 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Bladder | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 1 | 0 |
| Skin | 4 | 7 | 7 | 3 | 2 | 7 | 1 | 3 | 2 | 1 | 1 | 4 | 0 | 0 | 0 | 6 | 0 | 1 | 0 | 3 | 4 | 2 | 2 |
| Pancreas | 4 | 5 | 5 | 2 | 1 | 5 | 2 | 5 | 2 | 0 | 0 | 4 | 0 | 0 | 0 | 4 | 0 | 0 | 3 | 0 | 4 | 1 | 2 |
| Spleen | 3 | 5 | 5 | 2 | 1 | 5 | 1 | 4 | 2 | 1 | 0 | 4 | 0 | 0 | 0 | 4 | 0 | 1 | 4 | 4 | 0 | 2 | 2 |
| Ocular | 1 | 3 | 3 | 1 | 0 | 3 | 0 | 2 | 0 | 1 | 0 | 3 | 0 | 0 | 0 | 3 | 0 | 1 | 2 | 1 | 2 | 0 | 0 |
| Breast | 1 | 2 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 2 | 2 | 0 | 0 |

(9)

## Appendix C. Filtered co-occurrence matrix $\mathbf{A_F}$

The filtered co-occurrence matrix $A_F$ (Eq. (10)) is the submatrix of the full $23 \times 23$ matrix $A$ (Eq. (9)), restricted to the 11 source organs in $V_S = \{\text{Liver, Kidney, Thyroid, Stomach, Colon, Rectum, Prostate, Lung, Ovary, Bladder, Pancreas}\}$. Thus, $A_F = A[V_S, V_S]$ is the adjacency matrix of the subgraph induced by $V_S$.

$$A_F = \begin{array}{c} \\ \text{Liver} \\ \text{Kidney} \\ \text{Thyroid} \\ \text{Stomach} \\ \text{Colon} \\ \text{Rectum} \\ \text{Prostate} \\ \text{Lung} \\ \text{Ovary} \\ \text{Bladder} \\ \text{Pancreas} \end{array} \begin{array}{c} \begin{array}{ccccccccccc} \text{Liver} & \text{Kidney} & \text{Thyroid} & \text{Stomach} & \text{Colon} & \text{Rectum} & \text{Prostate} & \text{Lung} & \text{Ovary} & \text{Bladder} & \text{Pancreas} \end{array} \\ \begin{bmatrix} 0 & 4 & 2 & 2 & 3 & 1 & 1 & 19 & 2 & 1 & 5 \\ 4 & 0 & 1 & 0 & 1 & 1 & 1 & 3 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 2 \\ 2 & 0 & 0 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & 0 \\ 3 & 1 & 0 & 1 & 0 & 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 19 & 3 & 1 & 2 & 3 & 1 & 1 & 0 & 2 & 1 & 4 \\ 2 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 5 & 2 & 2 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 \end{bmatrix} \end{array} \tag{10}$$

## Appendix D. Algorithmic implementation

This appendix presents the pseudocodes for the three algorithmic procedures that operationalise the validation framework. These methods bridge the theoretical null models with the statistical significance tests in the Results section. Specifically, Algorithm 1 defines the overarching validation pipeline, Algorithm 2 details the degree-preserving MCMC protocol for the configuration model, and Algorithm 3 outlines the helper functions for mobility and normalisation. Together, these pseudocodes illustrate the study's logical architecture, while the executable MATLAB source code and datasets are available in the repository [57].

### *D.1. Metastatic network validation pipeline*

See Algorithm 1.

### *D.2. Degree-preserving MCMC with Metropolis–Hastings*

See Algorithm 2.

### *D.3. Helper functions*

See Algorithm 3.

## Data availability

The MATLAB implementation of the hypergraph framework (including the configuration null models, functional partitioning algorithms, and visualisation scripts) is open-source and permanently archived at Zenodo. The software can be accessed and cited via the following persistent identifier: https://doi.org/10.5281/zenodo.17542956 ([57]).

**Algorithm 1:** Metastatic Network Validation

**Input:** Incidence matrix $H \in \{0,1\}^{M\times N}$, source organs $V_S$, simulations $\mathcal{N}_{\text{sim}} = 10^4$
**Output:** $p$-value matrices $p^{\text{ER}}$, $p^{\text{conf}}$
**Step 1: Hypergraph Construction**;
$H \leftarrow$ load incidence matrix (Appendix A) Appendix **Step 2: Network Projection**;
$D_v \leftarrow \text{diag}(d(v_1), \ldots, d(v_M))$ ; // Vertex degrees
$A \leftarrow H H^T - D_v$ ; // Organ--organ co-occurrence
**Step 3:** $(V_S, V_T)$ **Partitioning**;
$A_F \leftarrow A[V_S, V_S]$ ; // Source organs only
**Step 4: Probabilistic Association Matrix**;
$\mathcal{S}_{mn} \leftarrow A_{F,mn} / \sum_{\ell \in V_S} A_{F,m\ell}$ ; // Conditional probabilities
**Step 5: Hierarchical Null Model Validation**;
*Erdős–Rényi Model:*;
$p^{\text{ER}} \leftarrow \mathbf{0}_{|V_S|\times|V_S|}$ ; // Initialise $p$-value matrix
**for** $\kappa \leftarrow 1$ **to** $\mathcal{N}_{sim}$ **do**
  // Random permutation preserving $\Lambda = \sum_{i=1}^{M}\sum_{j=1}^{N} H_{ij}$
  $H_{\text{rand}} \leftarrow \text{randomPermutation}(H)$
  // Identical projection pipeline
  $D_{v,\text{rand}} \leftarrow \text{diag}(\text{sum}(H_{\text{rand}}, 2))$;
  $A_{\text{rand}} \leftarrow H_{\text{rand}} H_{\text{rand}}^T - D_{v,\text{rand}}$;
  $A_{F,\text{rand}} \leftarrow A_{\text{rand}}[V_S, V_S]$;
  $\mathcal{S}_{\text{rand}} \leftarrow \text{rowNormalize}(A_{F,\text{rand}})$;
  $p^{\text{ER}} \leftarrow p^{\text{ER}} + \mathbb{I}(\mathcal{S}_{\text{rand}} \geq \mathcal{S})$; // One-tailed test
**end**
$p^{\text{ER}} \leftarrow \frac{p^{\text{ER}}}{\mathcal{N}_{\text{sim}}}$;
*Configuration Model:*;
$p^{\text{conf}} \leftarrow \mathbf{0}_{|V_S|\times|V_S|}$ ; // Initialise $p$-value matrix
$H_{\text{curr}} \leftarrow \text{DegreePreservingMCMC}(H, 10^6)$ ; // Burn-in to reach stationarity
**for** $\kappa \leftarrow 1$ **to** $\mathcal{N}_{sim}$ **do**
  $H_{\text{curr}} \leftarrow \text{DegreePreservingMCMC}(H_{\text{curr}}, 10^4)$ ; // Thinning interval for independence
  // Identical projection pipeline
  $D_{v,\text{rand}} \leftarrow \text{diag}(\text{sum}(H_{\text{curr}}, 2))$;
  $A_{\text{rand}} \leftarrow H_{\text{curr}} H_{\text{curr}}^T - D_{v,\text{rand}}$;
  $A_{F,\text{rand}} \leftarrow A_{\text{rand}}[V_S, V_S]$;
  $\mathcal{S}_{\text{rand}} \leftarrow \text{rowNormalize}(A_{F,\text{rand}})$;
  $p^{\text{conf}} \leftarrow p^{\text{conf}} + \mathbb{I}(\mathcal{S}_{\text{rand}} \geq \mathcal{S})$; // One-tailed test
**end**
$p^{\text{conf}} \leftarrow \frac{p^{\text{conf}}}{\mathcal{N}_{\text{sim}}}$;

**Algorithm 2:** DegreePreservingMCMC($H$, $\Gamma$)

**Input:** Original incidence matrix $H$, number of steps $\Gamma$
**Output:** Randomised matrix $H_{\text{rand}}$ with preserved degrees
$H_{\text{current}} \leftarrow H$;
$\Omega(H_{\text{current}}) \leftarrow$ calculateMobility($H_{\text{current}}$) ; // Initial full calculation
**for** $t \leftarrow 1$ **to** $\Gamma$ **do**
  // Select two distinct rows (organs) and two distinct columns (primary cancers)
  $[i, i'] \leftarrow$ selectDistinctIndices($M$) ; // $v_i, v_{i'} \in V$
  $[j, j'] \leftarrow$ selectDistinctIndices($N$) ; // $e_j, e_{j'} \in E$
  // Check valid swap pattern: alternating 0-1 or 1-0
  **if** $(H_{ij} = 1 \wedge H_{ij'} = 0 \wedge H_{i'j} = 0 \wedge H_{i'j'} = 1)$ **then**
    $H_{\text{proposed}} \leftarrow H_{\text{current}}$;
    $H_{\text{proposed}}[i, j] \leftarrow 0$, $H_{\text{proposed}}[i, j'] \leftarrow 1$;
    $H_{\text{proposed}}[i', j] \leftarrow 1$, $H_{\text{proposed}}[i', j'] \leftarrow 0$;
    $\Omega(H_{\text{proposed}}) \leftarrow$ updateMobility($H_{\text{current}}, i, i', j, j'$) ; // Local update for efficiency
    $\alpha \leftarrow \min\left(1, \frac{\Omega(H_{\text{current}})}{\Omega(H_{\text{proposed}})}\right)$ ; // Metropolis--Hastings
    **if** *random*() $< \alpha$ **then**
      $H_{\text{current}} \leftarrow H_{\text{proposed}}$;
      $\Omega(H_{\text{current}}) \leftarrow \Omega(H_{\text{proposed}})$;
    **end**
  **end**
  **else if** $(H_{ij} = 0 \wedge H_{ij'} = 1 \wedge H_{i'j} = 1 \wedge H_{i'j'} = 0)$ **then**
    $H_{\text{proposed}} \leftarrow H_{\text{current}}$;
    $H_{\text{proposed}}[i, j] \leftarrow 1$, $H_{\text{proposed}}[i, j'] \leftarrow 0$;
    $H_{\text{proposed}}[i', j] \leftarrow 0$, $H_{\text{proposed}}[i', j'] \leftarrow 1$;
    $\Omega(H_{\text{proposed}}) \leftarrow$ updateMobility($H_{\text{current}}, i, i', j, j'$) ; // Local update for efficiency
    $\alpha \leftarrow \min\left(1, \frac{\Omega(H_{\text{current}})}{\Omega(H_{\text{proposed}})}\right)$ ; // Metropolis--Hastings
    **if** *random*() $< \alpha$ **then**
      $H_{\text{current}} \leftarrow H_{\text{proposed}}$;
      $\Omega(H_{\text{current}}) \leftarrow \Omega(H_{\text{proposed}})$;
    **end**
  **end**
**end**
**return** $H_{\text{current}}$

**Algorithm 3:** Helper Functions

**Function** *calculateMobility(H)*:
- $\Omega \leftarrow 0$;
- **for** $i \leftarrow 1$ **to** $M-1$ **do**
  - **for** $i' \leftarrow i+1$ **to** $M$ **do**
    - **for** $j \leftarrow 1$ **to** $N-1$ **do**
      - **for** $j' \leftarrow j+1$ **to** $N$ **do**
        - **if** $(H_{ij}=1 \wedge H_{ij'}=0 \wedge H_{i'j}=0 \wedge H_{i'j'}=1)$ **then**
          - $\Omega \leftarrow \Omega+1$;
        - **end**
        - **if** $(H_{ij}=0 \wedge H_{ij'}=1 \wedge H_{i'j}=1 \wedge H_{i'j'}=0)$ **then**
          - $\Omega \leftarrow \Omega+1$;
        - **end**
      - **end**
    - **end**
  - **end**
- **end**
- **return** $\Omega$; // Number of possible edge swaps

**Function** *rowNormalize($A_F$)*:
- $\mathcal{S} \leftarrow \mathbf{0}_{|V_S|\times|V_S|}$;
- **foreach** $v_m \in V_S$ **do**
  - denominator $\leftarrow \sum_{v_\ell \in V_S} A_{F,m\ell}$;
  - **if** *denominator* $> 0$ **then**
    - $\mathcal{S}_{m,:} \leftarrow A_{F,m,:}$/denominator;
  - **end**
- **end**
- **return** $\mathcal{S}$; // Conditional probability matrix

**Function** *selectDistinctIndices(max)*:
- $i_1 \leftarrow$ randomInteger$(1, \max)$;
- **repeat**
  - $i_2 \leftarrow$ randomInteger$(1, \max)$;
- **until** $i_2 \neq i_1$;
- **return** $i_1, i_2$;

## References


[1] Guerra A, Betancourt-Mar J, Llanos-Pérez J, Mansilla R, Nieto-Villar J. Metastasis models: Thermodynamics and complexity. In: Systems biology. Springer; 2023, p. 45–75. http://dx.doi.org/10.1007/978-1-0716-3577-3_4.

[2] Kiri S, Ryba T. Cancer, metastasis, and the epigenome. Mol Cancer 2024;23(1):154. http://dx.doi.org/10.1186/s12943-024-02069-w.

[3] Mani K, Deng D, Lin C, Wang M, Hsu ML, Zaorsky NG. Causes of death among people living with metastatic cancer. Nat Commun 2024;15(1):1519. http://dx.doi.org/10.1038/s41467-024-45307-x.

[4] Dunbar KJ, Efe G, Cunningham K, Esquea E, Navaridas R, Rustgi AK. Regulation of metastatic organotropism. Trends Cancer 2025;11(3):216–31. http://dx.doi.org/10.1016/j.trecan.2024.11.012.

[5] Massagué J, Obenauf AC. Metastatic colonization by circulating tumour cells. Nature 2016;529(7586):298–306. http://dx.doi.org/10.1038/nature17038.

[6] Paget S. The distribution of secondary growths in cancer of the breast. Lancet 1889;133(3421):571–3. http://dx.doi.org/10.1016/S0140-6736(00)49915-0.

[7] Li Y, Liu F, Cai Q, Deng L, Ouyang Q, Zhang XH-F, Zheng J. Invasion and metastasis in cancer: molecular insights and therapeutic targets. Signal Transduct Target Ther 2025;10(1):57. http://dx.doi.org/10.1038/s41392-025-02148-4.

[8] Fares J, Fares MY, Khachfe HH, Salhab HA, Fares Y. Molecular principles of metastasis: a hallmark of cancer revisited. Signal Transduct Target Ther 2020;5(1):28. http://dx.doi.org/10.1038/s41392-020-0134-x.

[9] Gao Y, Bado I, Wang H, Zhang W, Rosen JM, Zhang XH-F. Metastasis organotropism: redefining the congenial soil. Dev Cell 2019;49(3):375–91. http://dx.doi.org/10.1016/j.devcel.2019.04.012.

[10] Agrawal A, Javanmardi Y, Watson SA, Serwinski B, Djordjevic B, Li W, Aref AR, Jenkins RW, Moeendarbary E. Mechanical signatures in cancer metastasis. NPJ Biol. Phys Mech 2025;2(1):3. http://dx.doi.org/10.1038/s44341-024-00007-x.

[11] Wang C, Luo D. The metabolic adaptation mechanism of metastatic organotropism. Exp Hematol Oncol 2021;10(1):30. http://dx.doi.org/10.1186/s40164-021-00223-4.

[12] Welch DR, Hurst DR. Defining the hallmarks of metastasis. Cancer Res 2019;79(12):3011–27. http://dx.doi.org/10.1158/0008-5472.CAN-19-0458.

[13] Bautista J, López-Cortés A. Deciphering organotropism reveals therapeutic targets in metastasis. Front Cell Dev Biol. 2025;13:1677481. http://dx.doi.org/10.3389/fcell.2025.1677481.

[14] Mets-Database. A database of metastatic patterns of spread (mets). 2026, http://mets.getthediagnosis.org/. (Accessed 22 January 2026).

[15] National Cancer Institute. Metastatic cancer: When cancer spreads. 2026, https://www.cancer.gov/types/metastatic-cancer.

[16] Majhi S, Perc M, Ghosh D. Dynamics on higher-order networks: A review. J R Soc Interface 2022;19(188). http://dx.doi.org/10.1098/rsif.2022.0043.

[17] Barabási A-L, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. Nature Rev Genet 2011;12(1):56–68. http://dx.doi.org/10.1038/nrg2918.

[18] Chen L, Blumm N, Christakis NA, Barabasi A, Deisboeck TS. Cancer metastasis networks and the prediction of progression patterns. Br J Cancer 2009;101(5):749–58. http://dx.doi.org/10.1038/sj.bjc.6605214.

[19] Su R, Zhu Y, Zou Q, Wei L. Distant metastasis identification based on optimized graph representation of gene interaction patterns. Brief Bioinform 2022;23(1):bbab468. http://dx.doi.org/10.1093/bib/bbab468.

[20] Gosak M, Milojević M, Duh M, Skok K, Perc M. Networks behind the morphology and structural design of living systems. Phys Life Rev 2022;41:1–21. http://dx.doi.org/10.1016/j.plrev.2022.03.001.

[21] Majhi S, Perc M, Ghosh D. Dynamics on higher-order networks: a review. J R Soc Interface 2022;19(188):20220043. http://dx.doi.org/10.1098/rsif.2022.0043.

[22] Majhi S, Ghosh S, Pal PK, Pal S, Pal TK, Ghosh D, Završnik J, Perc M. Patterns of neuronal synchrony in higher-order networks. Phys Life Rev 2025;52:144–70. http://dx.doi.org/10.1016/j.plrev.2024.12.013.

[23] Anwar MS, Ghosh D. Neuronal synchronization in time-varying higher-order networks. Chaos 2023;33(7):073111. http://dx.doi.org/10.1063/5.0152942.

[24] Gao Y, Ji S, Han X, Dai Q. Hypergraph computation. Engineering 2024;40:188–201. http://dx.doi.org/10.1016/j.eng.2024.04.017.

[25] Battiston F, Cencetti G, Iacopini I, Latora V, Lucas M, Patania A, Young J-G, Petri G. Networks beyond pairwise interactions: Structure and dynamics. Phys Rep 2020;874:1–92. http://dx.doi.org/10.1016/j.physrep.2020.05.004.

[26] Estrada E, Rodriguez-Velazquez JA. Subgraph centrality in complex networks. Phys Rev E 2005;71(5):056103. http://dx.doi.org/10.1103/PhysRevE.71.056103.

[27] Fosdick BK, Larremore DB, Nishimura J, Ugander J. Configuring random graph models with fixed degree sequences. Siam Rev 2018;60(2):315–55. http://dx.doi.org/10.1137/16M1087175.

[28] Margarit DH, Reale MV, Paccosi G, Romanelli LM. Decoding cancer dynamics: Hypergraph analysis of stem cell marker interactions through O-information and community detection. Phys Rev E 2025;111(5):054410. http://dx.doi.org/10.1103/PhysRevE.111.054410.

[29] Margarit DH, Paccosi G, Reale MV, Romanelli LM. Unveiling cancer stem cell marker networks: A hypergraph approach. Comput Biol Chem 2025;108684. http://dx.doi.org/10.1016/j.compbiolchem.2025.108684.

[30] International Agency for Research on Cancer (IARC). Global cancer burden growing, amidst mounting need for services. 2026, https://gco.iarc.fr/en.

[31] World Cancer Research Fund. Worldwide cancer data. 2026, https://www.wcrf.org/preventing-cancer/cancer-statistics/worldwide-cancer-data/.

[32] Global Burden of Disease. Worldwide cancer data. 2026, https://www.healthdata.org/data-tools-practices/interactive-data-visuals/gbd-cancer-compare.

[33] Allart-Vorelli P, Porro B, Baguet F, Michel A, Cousson-Gélie F. Haematological cancer and quality of life: a systematic literature review. Blood Cancer J 2015;5(4):e305. http://dx.doi.org/10.1038/bcj.2015.29.

[34] Hui JY. Epidemiology and etiology of sarcomas. Surg Clin 2016;96(5):901–14. http://dx.doi.org/10.1016/j.suc.2016.05.005.

[35] Tasci E, Zhuge Y, Camphausen K, Krauze AV. Bias and class imbalance in oncologic data—towards inclusive and transferrable AI in large scale oncology data sets. Cancers 2022;14(12):2897. http://dx.doi.org/10.3390/cancers14122897.

[36] Berge C. Hypergraphs: Combinatorics of Finite Sets, vol. 45, Elsevier; 1984.

[37] Gerstberger S, Jiang Q, Ganesh K. Metastasis. Cell 2023;186(8):1564–79. http://dx.doi.org/10.1016/j.cell.2023.03.003.

[38] Horn SR, Stoltzfus KC, Lehrer EJ, Dawson LA, Tchelebi L, Gusani NJ, Sharma NK, Chen H, Trifiletti DM, Zaorsky NG. Epidemiology of liver metastases. Cancer Epidemiol. 2020;67:101760. http://dx.doi.org/10.1016/j.canep.2020.101760.

[39] Dudani S, De Velasco G, Wells JC, Gan CL, Donskov F, Porta C, Fraccon A, Pasini F, Lee JL, Hansen A, et al. Evaluation of clear cell, papillary, and chromophobe renal cell carcinoma metastasis sites and association with survival. JAMA Netw Open 2021;4(1):e2021869. http://dx.doi.org/10.1001/jamanetworkopen.2020.21869.

[40] Nooreldeen R, Bach H. Current and future development in lung cancer diagnosis. Int J Mol Sci 2021;22(16):8661. http://dx.doi.org/10.3390/ijms22168661.

[41] Jayakrishnan T, Ng K. Early-onset gastrointestinal cancers: a review. Jama 2025;334(15):1373–85. http://dx.doi.org/10.1001/jama.2025.10218.

[42] Ben-Aharon I, van Laarhoven HW, Fontana E, Obermannova R, Nilsson M, Lordick F. Early-onset cancer in the gastrointestinal tract is on the rise—evidence and implications. Cancer Discov 2023;13(3):538–51. http://dx.doi.org/10.1158/2159-8290.CD-22-1038.

[43] Rajabi S, Shakib H, Dastmalchi R, Danesh-Afrooz A, Karima S, Hedayati M. Metastatic propagation of thyroid cancer; organ tropism and major modulators. Future Oncol 2020;16(18):1301–19. http://dx.doi.org/10.2217/fon-2019-0780.

[44] Mohamed OA, Tesen HS, Hany M, Sherif A, Abdelwahab MM, Elnaggar MH. The role of hypoxia on prostate cancer progression and metastasis. Mol Biol Rep 2023;50(4):3873–84. http://dx.doi.org/10.1007/s11033-023-08251-5.

[45] Dhaliwal D, Shepherd TG. Molecular and cellular mechanisms controlling integrin-mediated cell adhesion and tumor progression in ovarian cancer metastasis: a review. Clin Exp Metastasis 2022;39(2):291–301. http://dx.doi.org/10.1007/s10585-021-10136-5.
[46] Hammouz RY, Kołat D, Kałuzińska Ż, Płuciennik E, Bednarek AK. MicroRNAs: their role in metastasis, angiogenesis, and the potential for biomarker utility in bladder carcinomas. Cancers 2021;13(4):891. http://dx.doi.org/10.3390/cancers13040891.
[47] Gillot L, Baudin L, Rouaud L, Kridelka F, Noël A. The pre-metastatic niche in lymph nodes: formation and characteristics. Cell Mol Life Sci 2021;78(16):5987–6002. http://dx.doi.org/10.1007/s00018-021-03873-z.
[48] Ryan C, Stoltzfus KC, Horn S, Chen H, Louie AV, Lehrer EJ, Trifiletti DM, Fox EJ, Abraham JA, Zaorsky NG. Epidemiology of bone metastases. Bone 2022;158:115783. http://dx.doi.org/10.1016/j.bone.2020.115783.
[49] Compérat E, Bardier-Dupas A, Camparo P, Capron F, Charlotte F. Splenic metastases: clinicopathologic presentation, differential diagnosis, and pathogenesis. Arch Pathol Lab Med 2007;131(6):965–9. http://dx.doi.org/10.5858/2007-131-965-SMCPDD.
[50] Roth L, Russo L, Ulugoel S, Freire dos Santos R, Breuer E, Gupta A, Lehmann K. Peritoneal metastasis: current status and treatment options. Cancers 2021;14(1):60. http://dx.doi.org/10.3390/cancers14010060.
[51] Strange CD, Ahuja J, Faiz SA, Grosu HB, Harding WC, Keshava K, Jimenez CA, Shannon VR, De Groot PM, Truong MT, et al. The Pleura. In: The thorax: medical, radiological, and pathological assessment. Springer; 2023, p. 3–99. http://dx.doi.org/10.1007/978-3-031-21040-2_1.
[52] Engelking M, Rao M. Management of diaphragm tumors. Thorac Surg Clin 2024;34(2):189–95. http://dx.doi.org/10.1016/j.thorsurg.2024.01.009.
[53] Ioannidis JP. What have we (not) learnt from millions of scientific papers with P values? Amer Statist 2019;73(sup1):20–5. http://dx.doi.org/10.1080/00031305.2018.1447512.
[54] Perneger TV. What's wrong with Bonferroni adjustments. Bmj 1998;316(7139):1236–8. http://dx.doi.org/10.1136/bmj.316.7139.1236.
[55] Erdős P, Rényi A. On random graphs I. Publ Math 1959;6(3–4):290–7.
[56] Coolen AC, De Martino A, Annibale A. Constrained Markovian dynamics of random graphs. J Stat Phys 2009;136(6):1035–67. http://dx.doi.org/10.1007/s10955-009-9821-2.
[57] Margarit DH. Supplementary material for "higher-order architecture of metastatic networks: A statistical mechanics approach via validated hypergraphs." V1. 2026, http://dx.doi.org/10.5281/zenodo.17542956.
[58] Popper HH. Progression and metastasis of lung cancer. Cancer Metastasis Rev 2016;35(1):75–91. http://dx.doi.org/10.1007/s10555-016-9618-0.
[59] Ruivo A, Oliveira RC, Silva-Vaz P, Tralhão JG. Colorectal cancer liver metastasis—state-of-the-art and future perspectives. Gastrointest Disord 2023;5(4):580–608. http://dx.doi.org/10.3390/gidisord5040046.