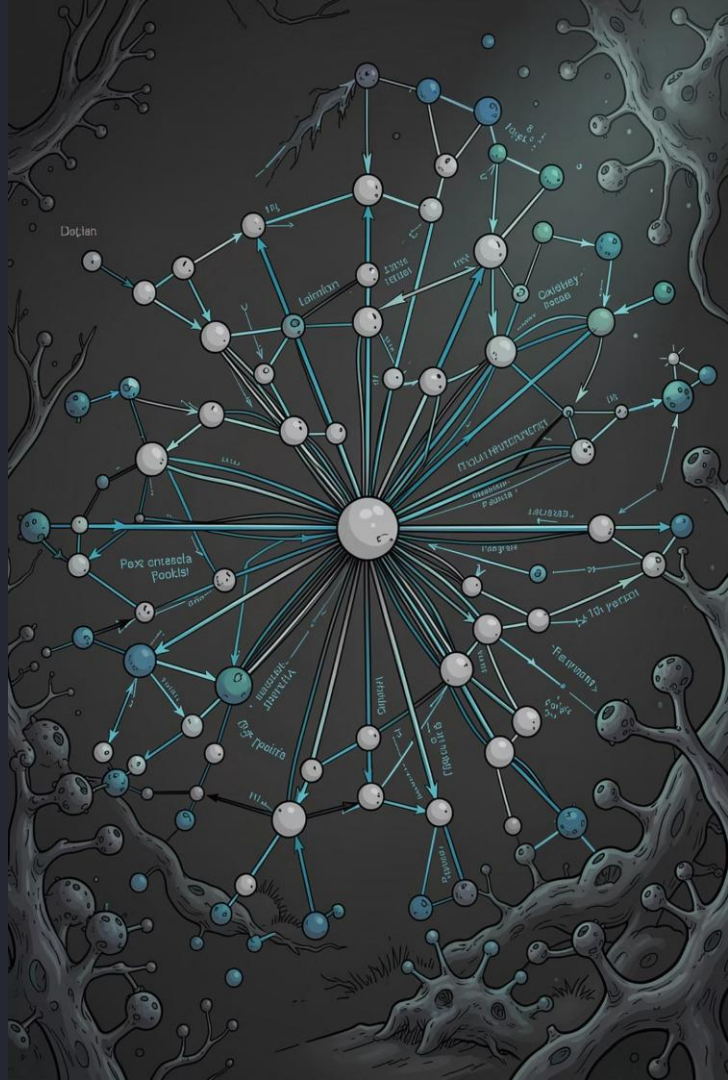
Mechanistic Network-Based Drug Repurposing for COVID-19

From Classical Barabási Network Proximity to INDRA-Based Mechanistic Proximity

Gyöngyi Kocsisné Szilágyi



Presentation Roadmap

Mechanistic Network-Based Drug Repurposing

1. Classical Barabási Network Proximity

- Protein–protein interaction (PPI) network
- Disease module and drug target sets
- Shortest-path proximity calculation
- Random baseline and Z-score framework

2. Extension to INDRA Mechanistic Networks

- Directed causal biological interactions
- Evidence-weighted mechanistic relations
- Drug–target mechanistic path extraction
- DrugBank-supported filtering

3. COVID Disease Module Construction

- Integration of DisGeNET, CTD, KEGG, and Disease Map
- Selection of 82 COVID-associated genes
- Construction of the final disease module

4. Mechanistic Drug Repurposing Workflow

- Extraction of drug–target paths
- Mechanistic proximity computation
- Random baseline generation
- Statistical significance analysis

5. Drug Candidate Identification

- Z-score distribution analysis
- Multi-metric drug ranking
- Top candidate prioritization
- Comparison of ranking strategies

6. Evidence-Based Clinical Evaluation

- Comparison with classical Barabási candidates
- Clinical trial and outcome analysis
- Positive ratio and enrichment evaluation
- Odds ratio and negative fraction analysis

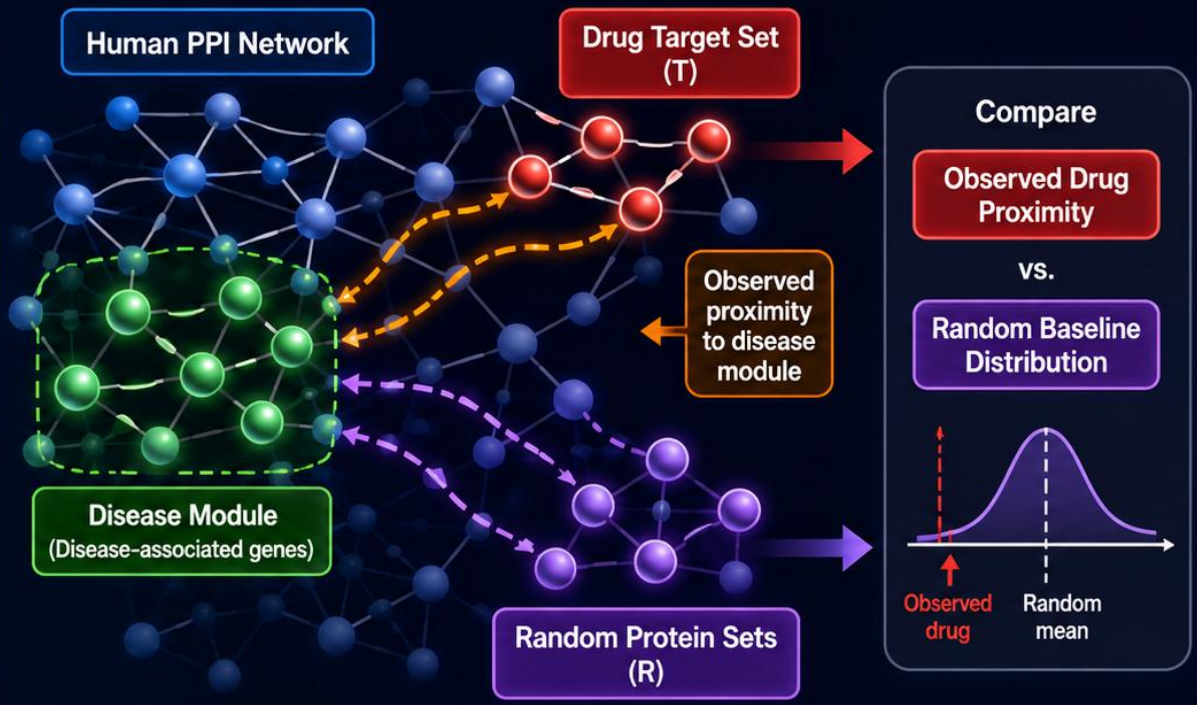
7. Conclusions and Future Directions

- Mechanistic vs. classical proximity
- Biological interpretation of candidate drugs
- Advantages and limitations of the framework
- Future validation possibilities

Classical Barabási Network Proximity

Measuring how close drug targets are to disease-associated genes in a protein–protein interaction network

- Start from a human protein–protein interaction (PPI) network.
- Define a disease module containing disease-associated genes.
- Represent each drug by its target protein set.
- Compute the shortest network distance between drug targets and the disease module.
- Generate random protein sets of the same size as the drug target set.
- Compare observed drug proximity against the random baseline.
- Drugs with significantly smaller distances may have therapeutic relevance.



The goal is to identify drugs whose targets are significantly closer to the disease module than expected by random chance.

Classical Barabási Network Proximity

A foundational framework for quantifying the topological closeness between drug targets and disease-associated genes in a human protein–protein interaction (PPI) network.

1. Disease Module

A collection of proteins and genes significantly associated with the disease phenotype.

2. Drug Target Set

The set of proteins directly targeted by a specific drug or compound.

3. Shortest Path Distance

The minimal network distance between a drug target protein (t) and its closest disease-associated gene (s).

4. Network Proximity (d(S,T))

Quantifies the average shortest path distance between drug targets (T) and disease genes (S).

PROXIMITY DEFINITION

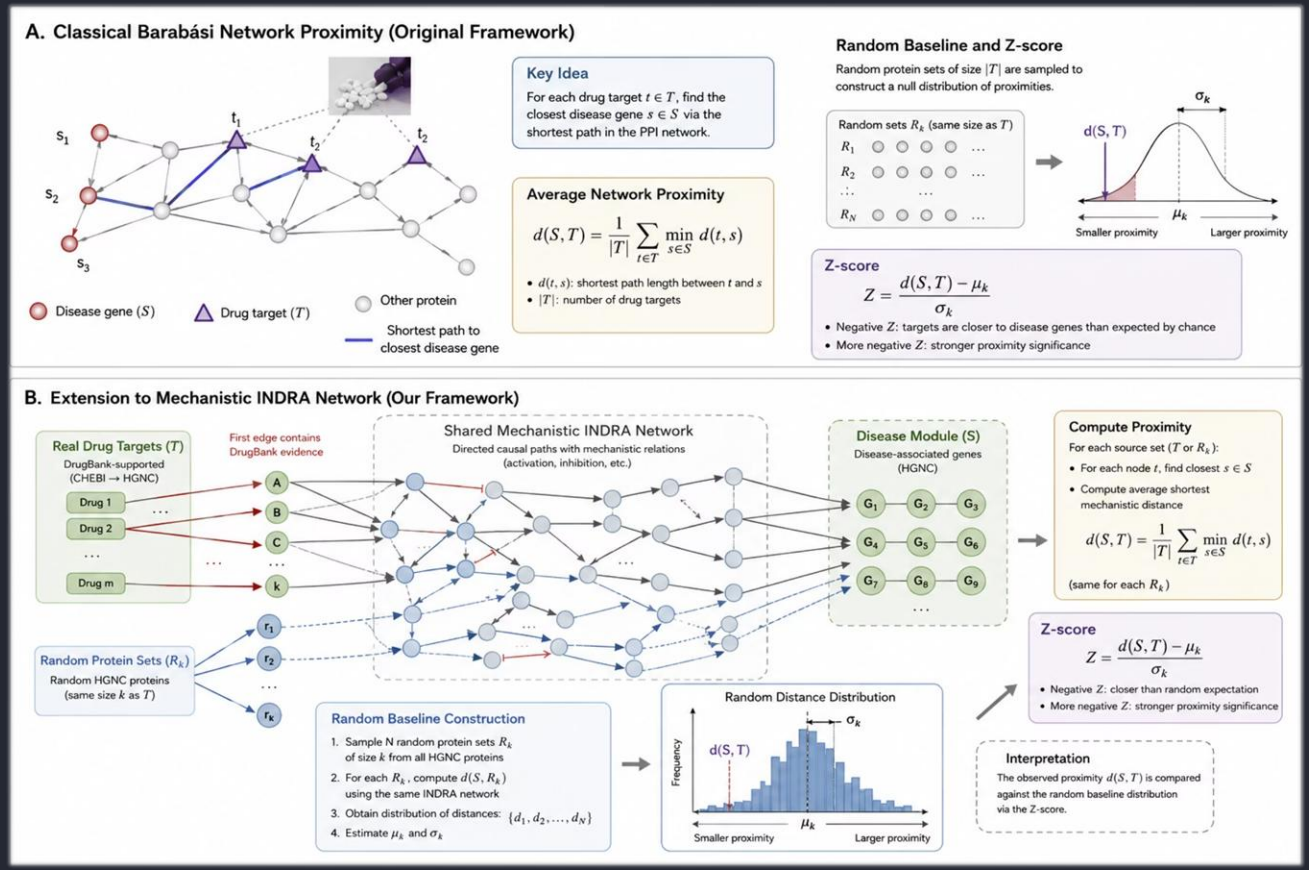
$$d(S,T) = \frac{1}{|T|} \sum_{t \in T} \min_{s \in S} d(t,s)$$

T: set of drug target proteins
S: set of disease-associated genes
$d(t,s)$: shortest path length between t and s
$|T|$: number of drug target proteins

INTERPRETATION

A smaller d(S,T) indicates that the drug targets are, on average, topologically closer to disease-associated genes in the PPI network, suggesting a higher potential for therapeutic relevance.

Conceptual Framework: From Classical to Mechanistic Proximity



Observed vs. Random Proximity

A Observed vs. Random Baseline Proximity

Observed Drug Proximity

Average closest distance from drug targets to disease module

$$d(S,T) = \frac{1}{|T|} \sum_{t \in T} \min_{s \in S} d(t,s)$$

- T: real drug targets (the target set of the given drug)
- S: disease module (set of disease-associated genes)

The random protein set R_k is generated such that $|T| = |R_k| = k$.
(i.e., the random set has the same size as the drug target set of this drug)

Random Baseline Proximity

"N" number of random sets for every "k"

For each fixed k:

- we generate N independent random protein sets: $R_1^k, R_2^k, \ldots, R_N^k$
- each set satisfies: $|R_i^k| = k$

Therefore

For each k:

- there are exactly N random samples
- each sample produces one value: $d_i^k = \frac{1}{k} \sum_{t \in R_i^k} \min_{s \in S} d(t,s)$

Random Baseline Construction

1. Sample N random protein sets R_k of size k from all HGNC proteins
2. For each R_k, compute $d(S, R_k)$ using the same INDRA network
3. Obtain distribution of distances: $\{d_1, d_2, \ldots, d_N\}$
4. Estimate μ_k and σ_k

Random Baseline Distribution and Z-score

B Random Baseline Distribution and Z-score

Random Baseline Distribution

$$random_baseline[k] = [d_1^k, d_2^k, \ldots, d_N^k]$$

Distribution of proximities from N random protein sets of size $k = |T|$.

For each k, we compute:

$$\mu_k = \frac{1}{N} \sum_{i=1}^{N} d_i^k$$

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (d_i^k - \mu_k)^2}$$

Z-score (computed for each drug)

$$Z_{drug} = \frac{d(S, T_{drug}) - \mu_k}{\sigma_k}$$

T_{drug}: target set of the given drug

$k = |T_{drug}|$

We compute Z_{drug} for every drug.

Negative Z_{drug}: closer than random expectation.

Sample Values (example)

k	μ_k	σ_k
1	1.7650	0.8943
2	1.6275	0.5338
3	1.6967	0.5155
4	1.7450	0.4750
5	1.7270	0.4070

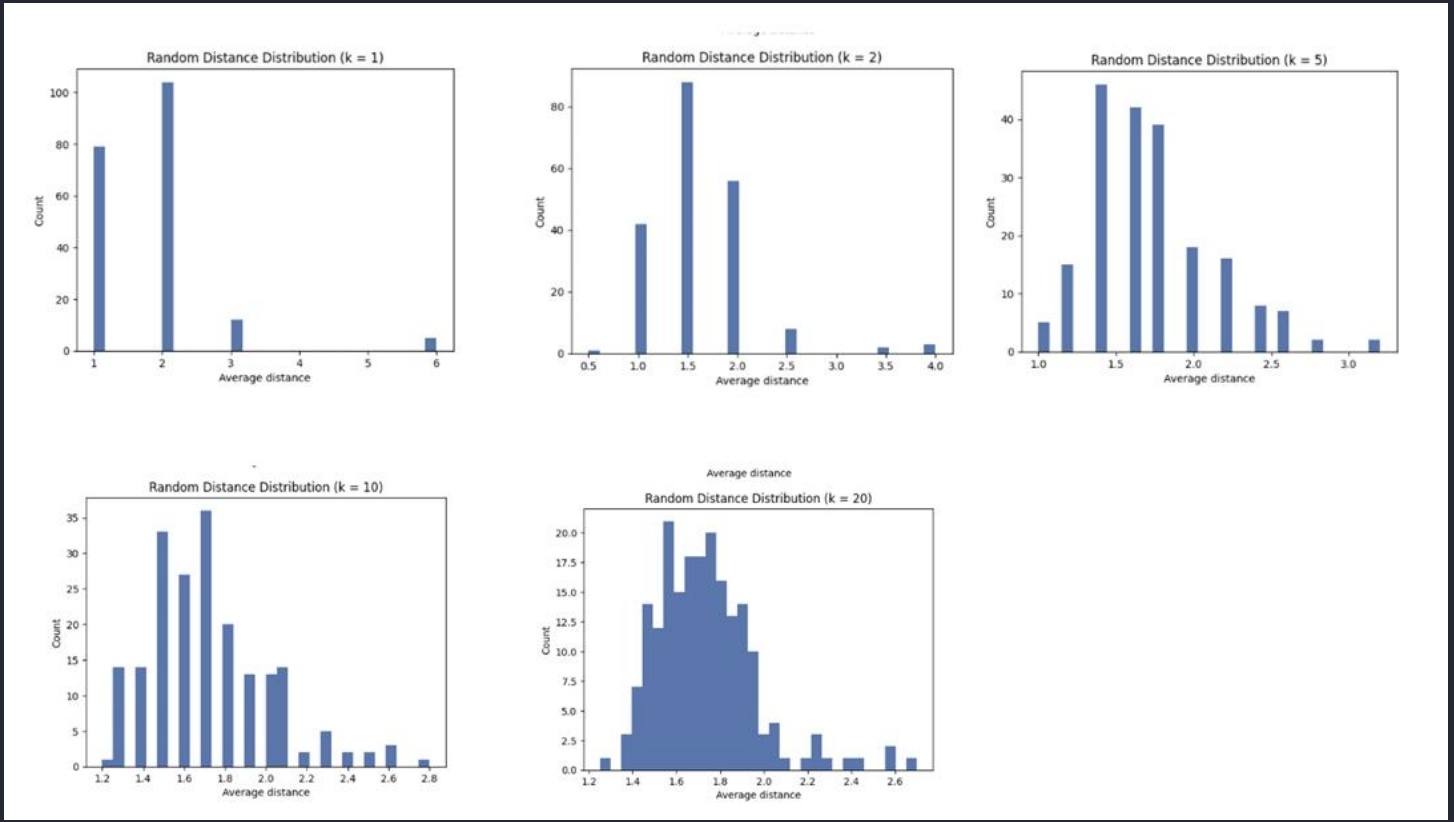
(Values are computed from N random samples for each k.)

Notation

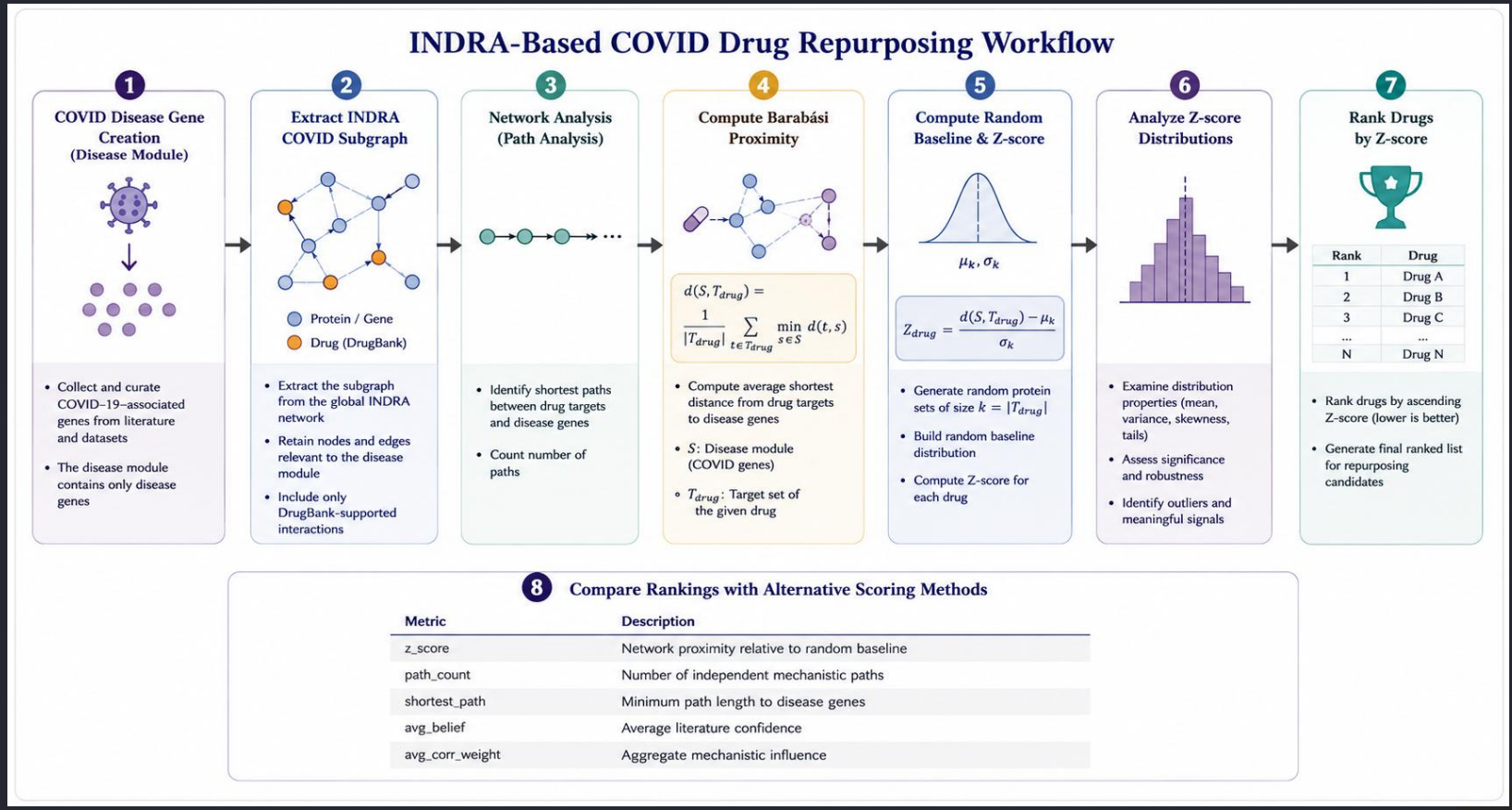
- S : disease module (set of disease-associated genes)
- T : real drug targets (target set of the given drug)
- R_i^k : i-th random protein set of size k
- $d(t, s)$: shortest path distance between proteins t and s in the PPI network
- $d(S, T)$: observed drug proximity
- d_i^k : proximity of i-th random set
- μ_k : mean of random proximities
- σ_k : standard deviation of random proximities
- Z_{drug} : Z-score for the given drug

Random Baseline Distribution

Random baseline distributions become progressively smoother and more Gaussian-like as k increases.



INDRA-Based COVID Drug Repurposing Workflow



1. COVID Disease Module Construction

DATA SOURCES

DisGeNET	10 genes	score ≥ 0.9	High-confidence clinically validated genes
CTD	508 genes	Inference Score ≥ 15; Reference Count ≥ 5	Curated disease-gene associations
KEGG hsa05171	82 genes	9 functional categories	Pathway-curated genes
COVID-19 Disease Map	46 genes	Interferon-1 SBML network	Mechanistic gene set

FINAL SELECTION RATIONALE

- Balances mechanistic specificity and biological coverage
- Avoids excessive noise from the FULL set
- Less restrictive than the CORE set
- High annotation coverage (96.3%)

CORE (High Confidence)

$$\text{CORE} = DiseaseMap \cap KEGG$$

- 39 genes
- Mechanistically validated + pathway-supported

EXTENDED (Mechanistic + Pathway)

$$\text{EXTENDED} = DiseaseMap \cup KEGG$$

- 82 genes
- Broader but still biologically structured

FULL (Discovery Set)

$$\text{FULL} = KEGG \cup DiseaseMap \cup CTD \cup DisGeNET$$

- 567 genes
- Maximum coverage

VALIDATED (Clinical Subset)

$$\text{VALIDATED} = FULL \cap DisGeNET$$

- 10 genes
- Clinically supported subset

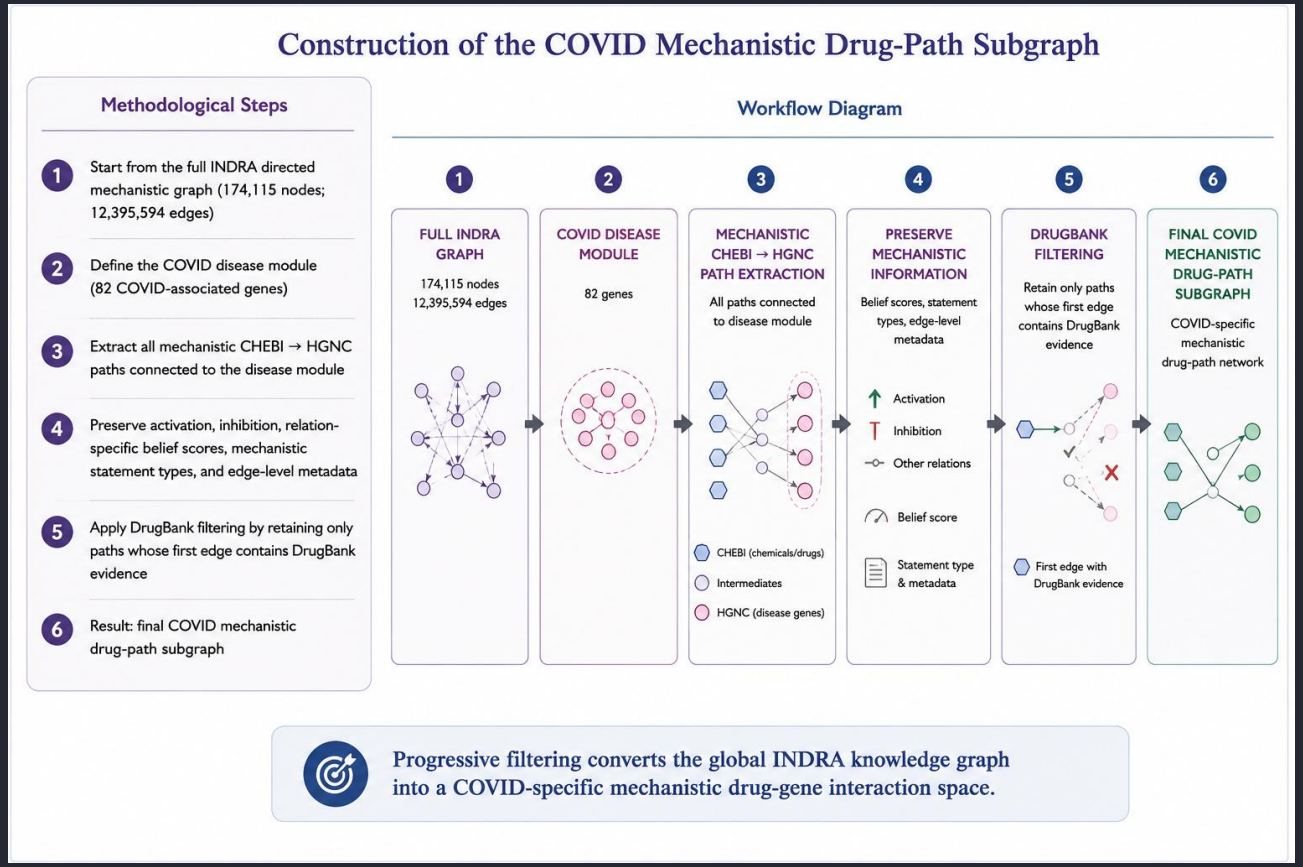
EXTENDED Gene Set (82 genes)

EXTENDED (Mechanistic + Pathway)

$$\text{EXTENDED} = DiseaseMap \cup KEGG$$

```
EXTENDED_GENES = set([
    "ACE","ACE2","ADAM17","AGTR1","C3","C5","C5AR1","CASP1","CCL2","CGAS",
    "CHUK","CSF2","CSF3","CXCL10","CXCL8","DDX58","EGFR","EIF2AK2","FOS","IFIH1",
    "IFNA1","IFNA2","IFNAR1","IFNAR2","IFNB1","IKBKB","IKBKE","IKBKG","IL1B","IL2",
    "IL6","IL6R","IL6ST","IRAK1","IRAK4","IRF3","IRF7","IRF9","ISG15","JAK1",
    "JUN","MAP3K7","MAPK1","MAPK14","MAPK3","MAPK8","MAS1","MAVS","MX1","MX2",
    "MYD88","NFKB1","NFKBIA","NLRP3","NRP1","OAS1","OAS2","OAS3","PYCARD","RELA",
    "RIPK1","STAT1","STAT2","STAT3","STING1","SYK","TAB1","TAB2","TANK","TBK1",
    "TICAM1","TLR2","TLR3","TLR4","TLR7","TLR8","TMPRSS2","TNF","TNFRSF1A","TRAF3",
    "TRAF6","TYK2"
])
```

2. Construction of the COVID-Specific Mechanistic Drug-Path Subgraph



3. INDRA COVID Subgraph Analysis (Path-Induced Network)

Subgraph Size and Structure

Metric	Value
Nodes	6841
Edges	14792

COVID Disease Module

Metric	Value
Disease genes	73

Top Hub Nodes (by Degree)

Node	Degree
ACE2	2464
IL6R	2423
MAPK3	1050
IL1B	916
NFKBIA	664
MAPK14	412
C5	373
TRAF6	366
STING1	343

Interpretation

- Hub nodes correspond to:
  - immune signaling (IL6R, IL1B, TNF)
  - viral entry (ACE2)
  - inflammation pathways (MAPK, NF-κB)
- Confirms biological relevance of the network

Nodes Most Connected to COVID Genes

Node	Connections
ACE2	2462
IL6R	2423
MAPK3	1045
IL1B	911
NFKBIA	661
MAPK14	394
C5	369
TRAF6	366
STING1	343
TNF	266

Interpretation

- Strong overlap with hub nodes
- Indicates that hubs are also **central mediators of disease connectivity**

Neighborhood of the Disease Module

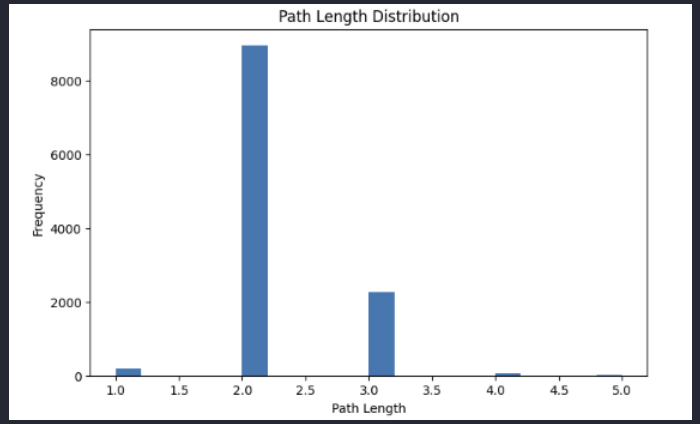
Metric	Value
Unique neighbors	6807

Interpretation

- Nearly the entire subgraph lies within the **extended neighborhood of disease genes**
- Indicates strong connectivity between drugs and disease module

Path Length Distribution

Path Length	Count
1	195
2	8945
3	2273
4	60
5	27



- Path lengths are strongly concentrated
- at **length 2** (8,945 paths),
- corresponding to the canonical sequence

  drug → target → disease gene.

4. Compute Barabási Proximity

Random baseline distributions become progressively smoother and more Gaussian-like as k increases.

Random Baseline Proximity

"N" number of random sets for every "k"

For each fixed k:

- we generate N independent random protein sets: $R_1^k, R_2^k, \ldots, R_N^k$
- each set satisfies: $|R_i^k| = k$

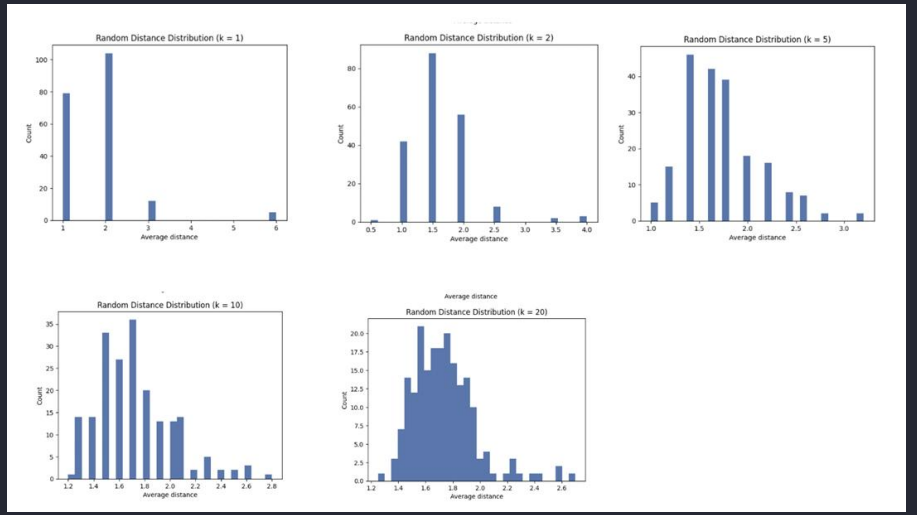
Therefore

For each k:

- there are exactly N random samples
- each sample produces one value: $d_i^k = \frac{1}{k} \sum_{t \in R_i^k} \min_{s \in S} d(t, s)$

$$random_baseline[k] = [d_1^k, d_2^k, \ldots, d_N^k]$$

Distribution of proximities from N random protein sets of size $k = |T|$.



5. Random Baseline Calculation

For each k,

we compute:

$$\mu_k = \frac{1}{N} \sum_{i=1}^{N} d_i^k$$

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (d_i^k - \mu_k)^2}$$

Baseline μ_k and σ_k Computation Results

Sample Values

k	μ_k	σ_k
1	1.7650	0.8943
2	1.6275	0.5338
3	1.6967	0.5155
4	1.7450	0.4750
5	1.7270	0.4070

Interpretation

- μ_k represents the expected average distance of k random proteins to the disease module
- σ_k measures the variability of this distance

6. Analyise Z-score Distribution

Goal

To analyze the statistical distribution of Barabási Z-scores computed for drug–disease proximity, and identify:

- overall distribution characteristics
- proportion of significant candidates
- strongest drug candidates

Method

The program performs:

1. **Statistical summary computation**
    - mean, standard deviation, percentiles
2. **Distribution analysis**
    - histogram
    - cumulative distribution
3. **Candidate identification**
    - negative vs positive Z-scores
    - strong candidates (Z < -2)
    - top-ranked drugs

6. Analyise Z-score Distribution

$$Z_{drug} = \frac{d(S, T_{drug}) - \mu_k}{\sigma_k}$$

T_{drug}: target set of the given drug

$k = |T_{drug}|$

We compute Z_{drug} for every drug.

Negative Z_{drug}: closer than random expectation.

Statistical Summary

Metric	Value
Count	3312
Mean	-0.8965
Std	0.7527
Min	-3.9749
25%	-1.1754
Median	-0.8554
75%	-0.8034
Max	1.7042

Key Counts

Category	Count
Negative Z	2804
Positive Z	508
Strong candidates (Z < -2)	204

Distance Scale Interpretation

d_{drug}	Meaning
≈ 0	direct target of disease gene
< 1	very close to disease module
≈ 1	one-step distance
> 1	more distant

Connection to Z-score

When:

$$d_{drug}(c) \ll \mu_k$$

then:

$$Z_c \ll 0$$

Implication

Drugs with:

- $d_{drug}(c) < 1$
- strongly negative Z_c

represent:

- high-confidence candidates
- direct mechanistic relevance

6. Analyise Z-score Distribution

Distribution Interpretation

1. Strong Negative Shift

- Mean Z ≈ **-0.90**
- Majority of values are **below zero**

Indicates that, on average, drugs are **closer to the disease module than random expectation**

2. Asymmetric Distribution

- Strong left tail (down to ~ -4)
- Limited positive range (up to ~ +1.7)

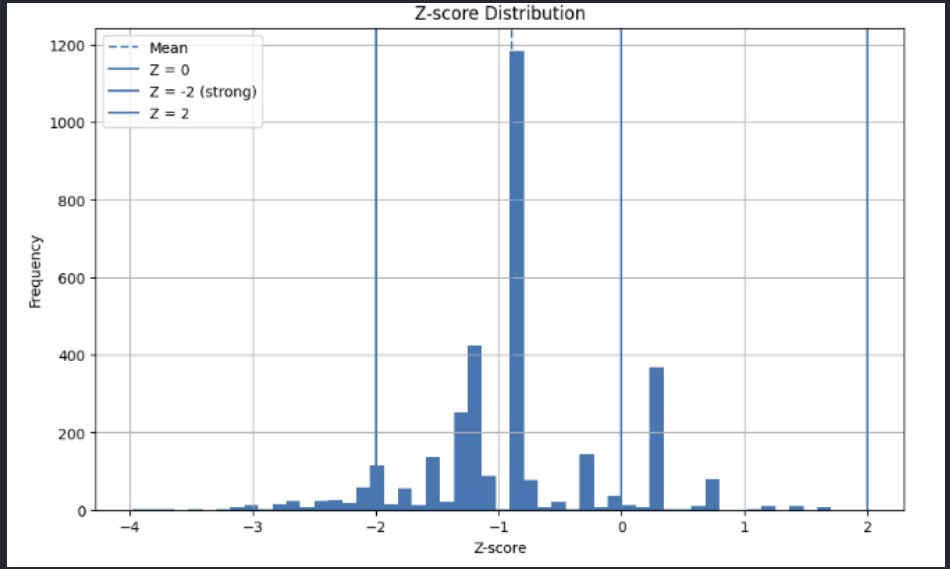
Suggests:

- presence of highly significant drugs
- relatively few strongly non-relevant drugs

3. High Proportion of Relevant Drugs

$$\frac{2804}{3312} \approx 84.6\%$$

Majority of drugs show **better-than-random proximity**



6. Analyise Z-score Distribution

Conclusion

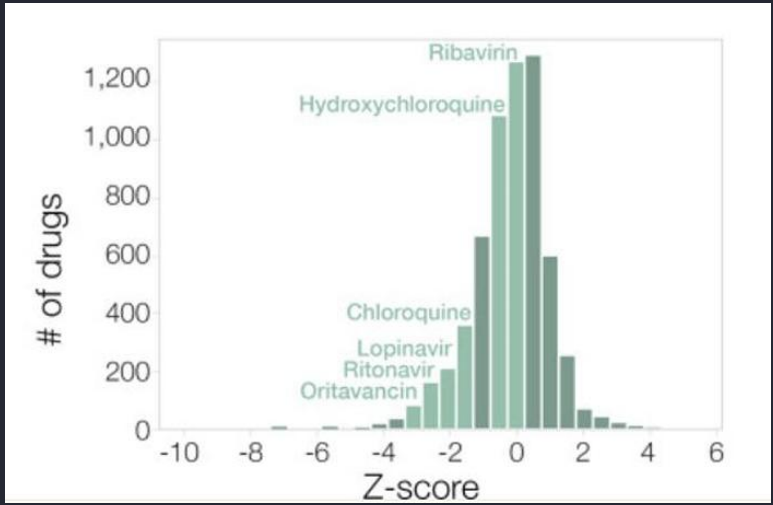
- PPI: symmetric distribution
- INDRA: **globally shifted distribution**

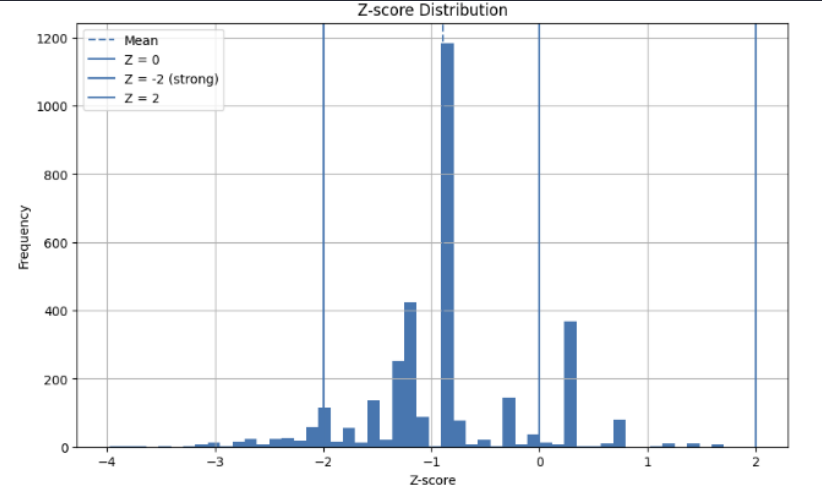
6.1 Evidence 1 - Gysi et al.

Network Medicine Framework for Identifying Drug Repurposing Opportunities for COVID-19

Deisy Morselli Gysi [1,2,*], Ítalo Do Valle [1,*], Marinka Zitnik [3,*], Asher Ameli [4,5,*], Xiao Gan [1,2,*], Onur Varol [1], Helia Sanchez [4], Rebecca Marlene Baron [6], Dina Ghiassian [4], Joseph Loscalzo [7], Albert-László Barabási [1,2,8]

„We tested the proximity of 6,116 drugs with at least one target in DrugBank, identifying 385 drugs with Z-Scores< −2, and 1,201 drugs with Z-Scores< −1, representing potential repurposing candidates (Fig. 4B).”





6. Analyise Z-score Distribution

1. Strong Tail (Z < -2)

PPI Evidence

- **Source:** Gysi et al. (COVID PPI repurposing study)
  https://pmc.ncbi.nlm.nih.gov/articles/PMC7280907.1/
- **Reported values:**
  - Total drugs: 6116
    "We tested the proximity of 6,116 drugs with at least one target in DrugBank, drugs with Z-Scores< –1, representing potential repurposing candidates "
  - Z < -2: 385
- **Derived:**
  - P(Z < -2) = 385 / 6116 ≈ **6.30%**

INDRA Result

- Z < -2 count: 204
- Total drugs: 3312
- **Computed:**
  - P(Z < -2) = 204 / 3312 ≈ **6.16%**

Conclusion

- Strong tail is **consistent across network types (~6%)**
- Indicates **similar extreme candidate detection**

2. Moderate Tail (Z < -1)

PPI Evidence

- **Source:** Gysi et al. (same study)
  https://pmc.ncbi.nlm.nih.gov/articles/PMC7280907.1/
- **Reported:**
  - Z < -1: 1201
- **Derived:**
  - P(Z < -1) = 1201 / 6116 ≈ **19.64%**

INDRA Result

- Z < -1 count: 1323
- **Computed:**
  - P(Z < -1) = 1323 / 3312 ≈ **39.95%**

Conclusion

- INDRA shows **~2× higher moderate proximity region**
- Indicates **broader enrichment**

6. Analyise Z-score Distribution

5. Dominance (P(Z < 0))

PPI Evidence

- **Source:** Guney / Gysi framework
  - https://pmc.ncbi.nlm.nih.gov/articles/PMC4740350/?utm_source=chatgpt.com
- **Not reported explicitly**, but:

given:

- symmetric distribution around 0
- **Derived assumption:**
  - $P(Z < 0) \approx \mathbf{0.5}$

INDRA Result

- $P(Z < 0) = \mathbf{0.8466}$
- $count(Z < 0) = 2804$

Conclusion

- INDRA shows **global dominance of proximity**
- ~85% of drugs are closer than random

Drug Candiadate Ranking by Z-Score

Top 20 Drug Candidates (Filtered)

Rank	Drug	Z-score
1	acetylsalicylic acid	-3.974866
2	pseudoephedrine	-3.953853
3	diarsenic trioxide	-3.805054
4	andrographolide	-3.751594
5	sulfasalazine	-3.664541
6	regorafenib	-3.445237
7	bromocriptine	-3.445237
8	haloperidol	-3.278428
9	Golotimod	-3.147543
10	talmapimod	-3.147543
11	polaprezinc	-3.141902
12	cabergoline	-3.092192
13	mianserin	-3.092192
14	bromhexine	-3.048620
15	imiquimod	-3.048620
16	resiquimod	-3.048620
17	Interferon beta-1b	-3.048620
18	Interferon alfa-2a	-3.048620
19	heparin	-2.997705
20	ponatinib	-2.997705

Drug Candiadate Comparison

Definitions Used in Outcome-Based Evaluation

Tested Candidate

A drug candidate was considered **tested** if it entered:

- a registered COVID-19 clinical trial
- therapeutic clinical investigation
- hospital treatment protocol
- or formal COVID-related observational therapeutic study

Untested or purely theoretical candidates were not included in the outcome statistics.

Outcome Categories

Positive

A candidate was classified as **positive** if:

- clinical studies demonstrated clear therapeutic benefit
- improvement in clinically relevant outcomes was reported
- or the compound became part of standard or supportive COVID treatment protocols

Examples:

- heparin
- corticosteroids

Drug Candiadate Comparison

Partial Positive

A candidate was classified as **partial positive** if:

- some studies reported beneficial effects
- positive effects were observed only in specific patient groups or treatment settings
- or evidence suggested moderate but non-universal clinical benefit

Examples:

- interferon beta-1b
- bromhexine
- aspirin

Drug Candiadate Comparison

Mixed / Inconclusive

A candidate was classified as **mixed/inconclusive** if:

- studies produced inconsistent results
- evidence remained insufficient
- or no clear consensus emerged regarding efficacy

Negative

A candidate was classified as **negative** if:

- large clinical trials failed to demonstrate meaningful therapeutic benefit
- or the candidate was later considered ineffective for COVID-19 treatment

Examples:

- hydroxychloroquine
- chloroquine
- lopinavir/ritonavir

Drug Candiadate Comparison

Evaluation Metrics

Positive Ratio

$$PR = \frac{N_{\text{positive}} + N_{\text{partial positive}}}{N_{\text{tested}}}$$

where:

- N_{tested} = number of clinically investigated candidates
- $N_{\text{positive}} + N_{\text{partial positive}}$ = number of candidates with fully or partially favorable COVID-related outcomes

Negative Fraction

$$NF = \frac{N_{\text{negative}}}{N_{\text{tested}}}$$

Enrichment Ratio

$$ER = \frac{PR_{\text{method}}}{PR_{\text{PPI}}}$$

Odds Ratio

$$OR = \frac{\left(\frac{N^{\text{method}}_{\text{positive}}}{N^{\text{method}}_{\text{nonpositive}}}\right)}{\left(\frac{N^{\text{PPI}}_{\text{positive}}}{N^{\text{PPI}}_{\text{nonpositive}}}\right)}$$

where the classical **PPI-based ranking** served as the reference framework.

Drug Candiadate Comparison

Evidence-Based COVID Outcome Table — Classical PPI / Barabási Candidates

Drug	Tested in COVID?	Evidence Source	Outcome Category	Evidence Basis / Rationale
Hydroxychloroquine	Yes	RECOVERY Trial, WHO Solidarity	Negative	No significant clinical benefit in large randomized trials
Chloroquine	Yes	WHO / multiple RCTs	Negative	Lack of efficacy and safety concerns
Lopinavir/Ritonavir	Yes	RECOVERY, WHO Solidarity	Negative	No meaningful mortality or recovery benefit
Ribavirin	Yes	Small combination trials	Mixed / Inconclusive	Limited and inconsistent antiviral evidence
Remdesivir	Yes	ACTT-1, FDA EUA	Partial Positive	Reduced recovery time in some hospitalized patients
Favipiravir	Yes	Multiple regional trials	Mixed / Inconclusive	Inconsistent efficacy across studies
Azithromycin	Yes	RECOVERY Trial	Negative	No meaningful COVID benefit
Losartan	Yes	Small RCTs	Mixed / Inconclusive	RAAS-modulation hypothesis not clearly validated
Rapamycin	Limited	Exploratory studies	Mixed / Inconclusive	Mechanistically interesting but insufficient evidence
Methylprednisolone	Yes	Corticosteroid clinical studies	Positive	Anti-inflammatory benefit in severe COVID

Drug Candiadate Comparison

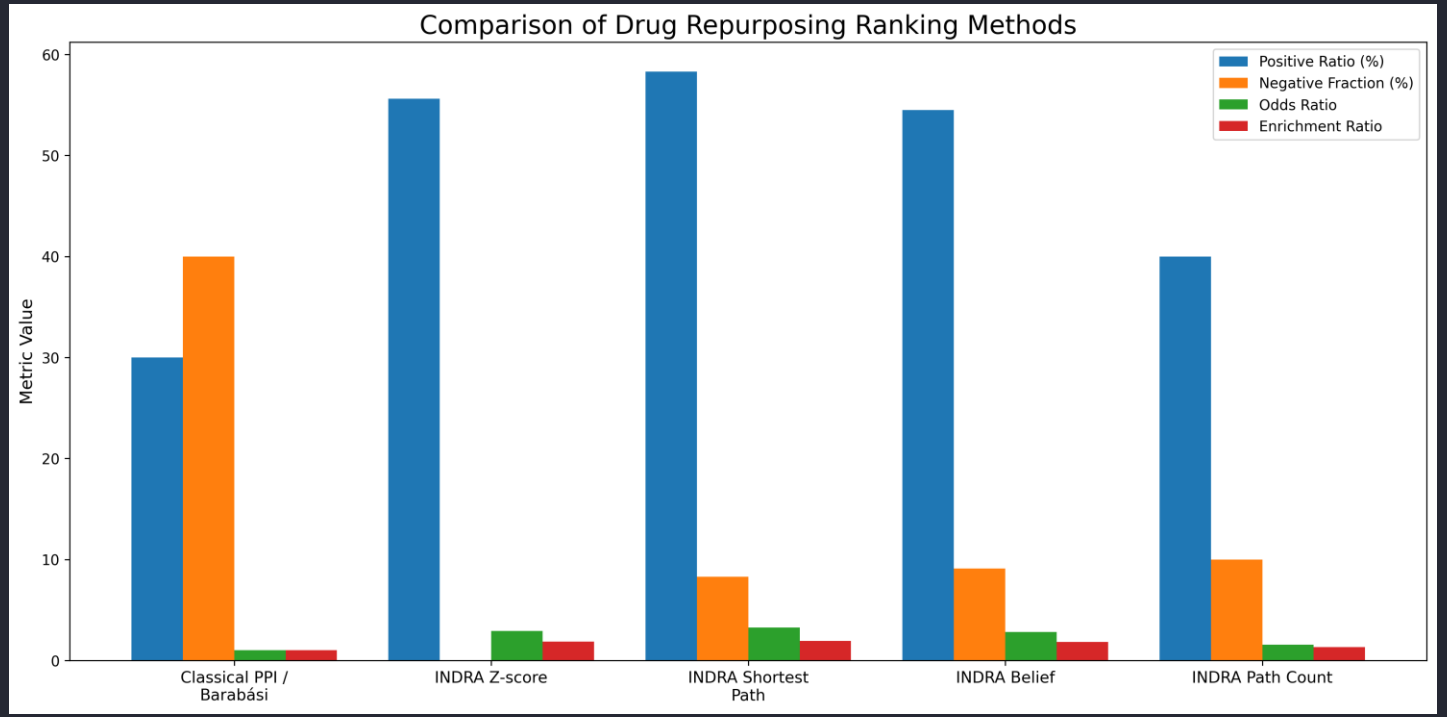
Evidence-Based COVID Outcome Table — INDRA Mechanistic Z-score Candidates

Drug	Tested in COVID?	Evidence Source	Outcome Category	Evidence Basis / Rationale
acetylsalicylic acid (aspirin)	Yes	Multiple observational and trial studies	Partial Positive	Antithrombotic and anti-inflammatory benefit
pseudoephedrine	No major COVID trials	No major evidence	Untested	Primarily symptomatic respiratory use
diarsenic trioxide	No	No COVID clinical evidence	Untested	Experimental mechanistic candidate only
andrographolide	Limited	Small clinical studies	Partial Positive	Preliminary antiviral and anti-inflammatory evidence
sulfasalazine	Limited	Small observational analyses	Mixed / Inconclusive	Immune modulation hypothesis
regorafenib	No	No COVID clinical evidence	Untested	Kinase signaling candidate only
bromocriptine	Limited	Small exploratory studies	Mixed / Inconclusive	Possible inflammatory modulation
haloperidol	Observational only	Retrospective studies	Mixed / Inconclusive	Neuroimmune observational hypotheses
Golotimod	No	No major evidence	Untested	Experimental immunomodulatory candidate
talmapimod	No	No major evidence	Untested	p38 MAPK inhibitor candidate
polaprezinc	No	No major evidence	Untested	Anti-inflammatory support hypothesis
cabergoline	No	No COVID evidence	Untested	Dopamine signaling hypothesis only
mianserin	No	No COVID evidence	Untested	Neuroimmune mechanistic hypothesis
bromhexine	Yes	Small clinical trials	Partial Positive	TMPRSS2 inhibition and respiratory benefit
imiquimod	Experimental	Preclinical / mechanistic studies	Untested	TLR7 immune activation candidate
resiquimod	Experimental	Preclinical immune studies	Untested	TLR7/8 agonist candidate
interferon beta-1b	Yes	Multiple RCTs	Partial Positive	Improved outcomes in some cohorts
interferon alfa-2a	Yes	Clinical investigations	Mixed / Partial Positive	Antiviral signaling relevance with mixed outcomes
heparin	Yes	HEP-COVID, guideline-supported	Positive	Reduced thrombotic complications and improved outcomes
ponatinib	Experimental	Preclinical only	Untested	Kinase inhibition hypothesis

Drug Candiadate Ranking

Ranking Method	Tested Candidates	Positive / Partial Positive	Clearly Negative	Positive Ratio	Negative Fraction	Odds Ratio (vs Classical PPI)	Enrichment Ratio
Classical PPI / Barabási	10	3	4	30.0%	40.0%	1.00	1.00
INDRA Z-score	9	5	0	55.6%	0.0%	2.92	1.85
INDRA Shortest Path	12	7	1	58.3%	8.3%	3.27	1.94
INDRA Belief	11	6	1	54.5%	9.1%	2.80	1.82
INDRA Path Count	10	4	1	40.0%	10.0%	1.56	1.33

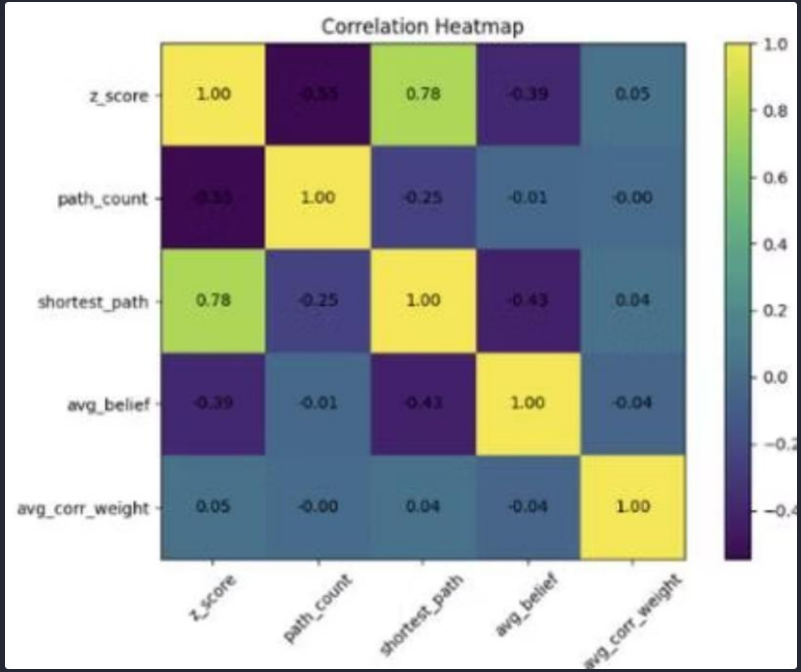
Drug Candiadate Comparison



Drug Candiadate Comparison

- Several important potential sources of bias should be considered when interpreting the results.
- **1. Temporal Knowledge Bias**
  The INDRA network likely contains more recent biological knowledge than the original 2020 Barabási framework.
  This does not mean that the graph "knew" the clinical outcomes, but rather that COVID-related mechanistic biology became more detailed over time.
- **2. Literature Density Bias**
  INDRA is literature-derived, meaning that heavily studied pathways such as **IL6**, **interferon**, **cytokine storm**, **thrombosis**, and **ACE2** may be overrepresented in the network.
- **3. DrugBank Edge Bias**
  Using only **DrugBank-supported first edges** improves specificity, but may favor well-known drugs and better-annotated targets.
- **4. Survivorship Bias**
  Negative clinical evidence is rarely represented explicitly in mechanistic biological networks, which may bias the graph toward positive mechanistic relations.
- **5. Validation Bias**
  The COVID clinical trial landscape itself was also biased by factors such as rapid repurposing efforts, media attention, drug availability, safety considerations, and existing FDA approval status.

Multi-Metric Drug Ranking



Correlation Analysis

Z-score ↔ Shortest Path: 0.78

Higher proximity is associated with shorter mechanistic path length.

Z-score ↔ Path Count: −0.55

Greater path multiplicity is associated with stronger proximity.

Z-score ↔ Belief: −0.39

Literature-derived belief alone shows a weaker association with proximity.

Z-score ↔ Corr. Weight: ~0.05

Edge weight shows negligible association with proximity.

Conclusions & Contributions

Disease module definition

Integrated gene sets from DisGeNET, CTD, KEGG, and Disease Map using hierarchical confidence tiers (CORE to FULL; 39–567 genes).

Mechanistic network model

INDRA represents directed causal relationships, capturing mechanistic structure not represented by undirected PPI networks.

Z-score analysis

Among 3,312 drugs, 85% had negative Z-scores; 204 met the strong-candidate threshold ($Z < -2$).

Prioritized candidates

Acetylsalicylic acid, pseudoephedrine, andrographolide, sulfasalazine, and bromocriptine were prioritized.

> DrugBank-derived targets were significantly closer to disease-associated genes in the INDRA mechanistic network than random protein sets.

