

# Knowledge Graph-Based Modeling of Compound–Target–Disease Associations for Drug Repurposing

Umesh A. Batra

Department of Computer Science, University of Central Florida, Orlando, FL, USA.

batra467@ucf.edu

Stefano Erickson

School of Computing, Clemson University, Clemson, SC, USA.

stefano.erickson@clemson.edu

## Abstract

Drug repurposing seeks to identify new therapeutic uses for existing compounds, yet the underlying evidence remains fragmented across chemical, biological, pharmacological, and clinical resources. Knowledge graphs provide a principled framework for integrating these heterogeneous relationships and reasoning over compound–target–disease associations in ways that isolated databases cannot support. This article presents a systems-level examination of knowledge graph-based modeling for drug repurposing, emphasizing architectural choices, governance mechanisms, inference trade-offs, and deployment constraints rather than algorithmic detail. The discussion addresses how entity alignment, ontology mapping, and provenance-aware ingestion shape the quality and trustworthiness of graph-based evidence. It further compares embedding-based and path-based inference, focusing on interpretability, scalability, robustness, and calibration under incomplete biomedical knowledge. The article also considers how graph-driven repurposing can influence research priorities and how governance structures can reduce disparities in representation across disease areas and populations. We argue that sustainable and responsible graph-based repurposing requires treating the knowledge graph as a governed socio-technical infrastructure rather than a static analytical artifact. Illustrative comparisons with network pharmacology and polypharmacy modeling are used to clarify structural trade-offs, and the discussion extends to data stewardship, model documentation, regulatory alignment, and institutional sustainability.

## Keywords

knowledge graphs, drug repurposing, compound-target-disease associations, systems pharmacology, biomedical informatics, graph machine learning, data governance.

## 1. Introduction

The declining productivity of de novo drug discovery has intensified interest in identifying new therapeutic uses for existing compounds, a strategy commonly known as drug repurposing. Repurposing offers important practical advantages, including existing pharmacokinetic and safety data, shorter development timelines, and reduced early-stage attrition. However, the evidence needed to support repurposing decisions is dispersed across many heterogeneous resources. DrugBank [1] provides curated information about drug targets and molecular mechanisms, but it represents only one part of a broader causal and associative landscape. A knowledge graph can integrate such chemical, biological, and clinical information into a unified multi-relational structure, enabling systematic exploration of indirect relationships among compounds, targets, and diseases. This article examines

knowledge graph-based modeling as an infrastructure for drug repurposing, focusing on its architectural foundations, governance challenges, inference properties, and policy implications.

The analytical core of graph-based drug repurposing is the recognition that therapeutic action is not usually a single molecular interaction but a path or subgraph involving multiple biological processes. Protein–protein interaction networks such as STRING [2] illustrate how drug targets are embedded in complexes, signaling cascades, and regulatory circuits. The Comparative Toxicogenomics Database [3] further contributes chemical–gene–disease relationships extracted from the literature, helping to connect compound exposure to phenotypic outcomes. When these resources are combined into a knowledge graph, a compound may be linked to a disease through multiple intermediate nodes, such as a primary target, an interacting protein, a pathway, a cellular process, and a phenotype. This representation is richer than a simple bipartite association matrix because it preserves the distinct semantic types of each edge and node.

Disease-oriented knowledge sources add another layer of explanatory depth. DisGeNET [4] integrates gene–disease associations from expert curation, text mining, and genome-wide studies. The Gene Ontology [5] provides a controlled vocabulary for biological processes, molecular functions, and cellular components, while the Human Phenotype Ontology [6] captures phenotypic abnormalities in a hierarchical form. These ontologies allow a knowledge graph to normalize different levels of biological granularity and support reasoning over shared mechanisms. Without such normalization, equivalent biological concepts may be represented inconsistently, producing fragmented graphs that weaken cross-domain inference. The systems perspective adopted here therefore treats ontology alignment and semantic standardization not as preprocessing details but as central governance decisions.

## **2. Background and Related Work**

Knowledge graphs have become a general-purpose framework for representing heterogeneous facts as typed entities and relations. Large collaborative platforms such as Wikidata [7] demonstrate how knowledge graphs can integrate diverse domains at scale while maintaining community-driven curation. Surveys of knowledge graph technologies [8] emphasize that graph-based systems differ from conventional relational databases because they treat the schema as an evolving set of semantic relationships rather than a fixed tabular structure. In biomedical applications, this flexibility is especially valuable because biological knowledge changes rapidly, and new entity types such as variants, metabolites, and microbiome components must be incorporated without disrupting existing data.

Learning from knowledge graphs has been advanced by embedding methods that project entities and relations into continuous vector spaces. Surveys of knowledge graph embedding [9] and relational machine learning [10] describe how such representations support link prediction, entity classification, and similarity search. Translation-based models such as TransE [11] introduced the influential idea of modeling a relation as a geometric translation between entity representations. More recent inductive methods have shifted attention from transductive entity embeddings to subgraph reasoning, allowing predictions for entities that were not present during training [12]. These developments are relevant to drug repurposing because new compounds, targets, and rare diseases continuously enter the biomedical literature.

Network-based drug repurposing has a substantial precedent. Hetionet [13] integrates multiple biomedical networks and has been used to prioritize drug–disease pairs based on network patterns. Reviews of drug repurposing describe both the promise and the translational barriers, including the difficulty of distinguishing causal signals from confounding associations [14]. Network pharmacology approaches have been reviewed extensively in the context of repositioning, emphasizing the need for robust community detection and path analysis [15]. The modeling of synergistic combinations through network communities has also been explored in herbal medicine and complex systems [16]. Although such work is not centered on conventional pharmaceutical development, it illustrates how community structure can reveal indirect therapeutic relationships.

Graph neural approaches have further demonstrated the utility of structured biological information. Polypharmacy side effect prediction based on multimodal graph convolution [17] shows that molecular and interaction data can be combined to anticipate adverse outcomes. Inductive representation learning on large graphs [18] has provided scalable mechanisms for generating node representations without retraining entire models, which is important for periodically updated biomedical knowledge graphs. Together, these lines of work establish the technical foundations for graph-based repurposing while also revealing unresolved challenges in governance, interpretability, and institutional deployment.

### **3. Knowledge Graph Construction and Schema Design for Compound–Target–Disease Associations**

The construction of a biomedical knowledge graph for drug repurposing requires careful decisions about identity, scope, and provenance. Identity resolution is fundamental because a compound may appear under different names, salt forms, and database identifiers across DrugBank [1], chemical registries, and clinical trial records. Similarly, a protein target may be recorded by gene symbol, UniProt accession, or sequence-derived identifier. Without robust entity resolution, the graph becomes fragmented, and a compound may fail to connect to its true target because the target is represented as two distinct nodes. The schema must therefore support cross-reference tables, canonical identifiers, and versioned mappings while preserving the original provenance of each assertion.

Ontology mapping constitutes a second architectural pillar. The Gene Ontology [5] and the Human Phenotype Ontology [6] provide semantic hierarchies that allow queries to expand from a specific phenotype to its ancestors, thereby capturing mechanistic relationships that would otherwise remain implicit. For example, a graph may link a target to a narrow cellular phenotype, while a disease is associated with a broader phenotypic class. Ontology-aware traversal can recover an indirect association between the target and the disease by navigating the hierarchical relation between phenotype terms. This kind of semantic closure is central to graph-based reasoning, but it also introduces risks if ontology versions change or if automated mappings are not validated against domain expertise.

Provenance and confidence are especially important in the biomedical domain because graph edges may originate from curated experiments, literature mining, sequence homology, or predictive models. A knowledge graph that collapses all evidence into a single unweighted edge cannot distinguish high-confidence causal interactions from speculative associations. Governance mechanisms should therefore attach provenance metadata, evidence type, source database, publication identifiers, and update timestamps to each edge. This design enables downstream ranking models to weight evidence by reliability and allows auditors to trace a prediction back to its supporting assertions. The systems implication is that the graph is not

merely a data structure but a continuously curated infrastructure with explicit epistemic commitments.

The scope of the graph also shapes what kinds of repurposing hypotheses can be generated. A minimal graph containing only compounds, targets, and diseases supports direct associative inference but cannot explain mechanism. Adding pathways, protein complexes, phenotypes, tissues, and orthologs increases explanatory depth while also increasing the complexity of schema maintenance. The structural trade-off is between breadth and semantic coherence. A broad graph assembled automatically from many sources may contain millions of edges but also substantial noise. A narrower graph curated by experts may be more reliable but may fail to surface novel connections in understudied areas. Therefore, schema design should be treated as a policy decision that reflects institutional priorities about recall, precision, and translational readiness.

#### **4. System Architecture and Infrastructure**

A production knowledge graph for drug repurposing requires an infrastructure that supports continuous ingestion, transformation, validation, and serving. Ingestion pipelines must connect to source databases such as DrugBank [1], STRING [2], DisGeNET [4], and ontology repositories, each of which has its own release cycle and data format. These pipelines should produce versioned graph snapshots so that any prediction can be reproduced against the exact state of the knowledge graph at a given time. Without versioning, changes in source data may silently alter model outputs and undermine scientific reproducibility. The infrastructure must therefore include data lineage tracking as a first-class system property rather than an afterthought.

Storage and query technologies introduce further architectural trade-offs. Graph databases support efficient traversal of entity–relation paths, while specialized graph computation engines can distribute iterative algorithms over very large networks. However, embedding-based inference often requires transforming the graph into tensors for training, which may necessitate separate storage for the symbolic graph and the learned representations. This separation can create operational friction because updates to the graph must be propagated into retrained embeddings. Systems that emphasize path-based reasoning may avoid frequent retraining but may incur higher query-time computational costs. The choice between materialized embeddings and online graph traversal is therefore not purely technical; it has direct consequences for latency, cost, interpretability, and maintenance effort.

Deployment of graph-based repurposing models also requires careful attention to update frequency and computational sustainability. Biomedical knowledge graphs grow rapidly, and periodic full retraining can be expensive in both energy and engineering time. Incremental representation learning methods [18] provide one route toward maintaining model utility under graph evolution. However, incremental updates may introduce conceptual drift if older assertions are retired or if semantics of ontology classes change. A sustainable infrastructure should balance freshness against stability, perhaps by maintaining multiple model versions and by defining release cadences that align with the rate of meaningful biomedical change rather than with arbitrary calendar intervals.

Security and access control are additional infrastructure concerns because drug repurposing workflows may involve proprietary compounds, unpublished clinical observations, or sensitive patient-derived data. The knowledge graph may need to support tiered access, in which public mechanistic knowledge is visible to all users while proprietary hypotheses and

validation results remain restricted. This requirement complicates governance because inference over a single graph may mix public and private edges, and model explanations may inadvertently reveal restricted information. System architects must therefore design access policies at the granularity of graph neighborhoods and derived predictions, not only at the level of entire datasets.

## **5. Inference and Ranking for Drug Repurposing**

Inference over a knowledge graph for drug repurposing generally falls into two broad families: path-based reasoning and embedding-based link prediction. Path-based methods exploit explicit sequences of relations connecting a compound to a disease, such as a compound inhibiting a target, the target participating in a pathway, and the pathway being implicated in a disease. These approaches are transparent because each prediction is accompanied by a concrete chain of evidence that can be inspected by domain experts. However, path-based reasoning can suffer from combinatorial explosion when many intermediate nodes are allowed, and it may miss meaningful patterns that involve multiple weak paths or graph motifs that are not captured by predefined path templates.

Embedding-based inference learns dense representations of entities and relations from the structure of the knowledge graph [9]. Translation-based models and other relational learning methods [10, 11] score candidate compound–disease pairs by comparing learned embeddings, allowing efficient ranking across very large candidate sets. These models can capture global statistical regularities that are not visible in any single path. However, their predictions are often difficult to explain in mechanistic terms, and they may overfit to highly connected nodes such as well-studied diseases and widely used compounds. The deployment of embedding-based systems therefore requires additional interpretability layers that map high-scoring predictions back to supporting graph neighborhoods.

A hybrid approach can combine the strengths of both families. Subgraph reasoning methods [12] generate predictions by considering local graph structure around a query, while graph neural models such as those applied to polypharmacy side effects [17] aggregate information from molecular and interaction graphs. In practice, a repurposing platform may use embedding models for candidate generation and subsequently apply path-based validation to produce mechanistic explanations for the most promising candidates. This staged design introduces a structural trade-off: candidate generation emphasizes recall and computational efficiency, whereas validation emphasizes precision and explanatory depth. The overall system design must explicitly manage this trade-off rather than assuming that a single inference algorithm can satisfy all translational requirements.

Uncertainty quantification is another major concern. Graph-based predictions are often treated as scores rather than probabilities, and the relationship between a high score and clinical promise is not straightforward. The incompleteness of the graph, the uneven density of evidence, and the presence of confounding associations all contribute to uncertainty. Reproducibility requires that inference outputs be accompanied by confidence intervals, sensitivity analyses, and evidence provenance. Reviews of network-based repositioning have highlighted that many published predictions fail to translate because they rely on convenient network metrics without rigorous validation [14, 15]. A systems-level design should therefore include built-in evaluation against held-out edges, temporal splits, and external experimental or clinical evidence.

## **6. Robustness, Fairness, and Governance**

Robustness in graph-based drug repurposing refers to the capacity of the system to maintain reliable predictions under incomplete, noisy, and evolving knowledge. Biomedical graphs are notoriously incomplete because many interactions have not been experimentally tested and many negative results are never published. Embedding models may be sensitive to this incompleteness, learning spurious regularities from the uneven sampling of biological space. Inductive methods that rely on local subgraph structure [12, 17] can be more robust to missing global statistics, but they may still inherit biases present in the underlying graph. Robustness should therefore be evaluated not only through aggregate link prediction accuracy but also through stress tests that simulate missing edges, deleted nodes, and shifted ontology versions.

Fairness is an increasingly important consideration in computational drug repurposing. Knowledge graphs are constructed from the scientific literature and curated databases, both of which are historically biased toward well-funded diseases and model organisms. Rare diseases, diseases predominantly affecting underserved populations, and neglected tropical diseases may be poorly represented in compound–target–disease knowledge graphs. As a result, graph-based ranking systems may systematically overlook repurposing opportunities for these conditions. Model cards and similar transparency documents [19] provide a mechanism for reporting the intended use, limitations, and evaluation context of predictive systems. When integrated into a repurposing platform, such documentation can make inequities visible and enable corrective action.

Data governance is inseparable from algorithmic governance. The FAIR principles for scientific data management [20] emphasize findability, accessibility, interoperability, and reusability, all of which are relevant to biomedical knowledge graphs. A FAIR-aligned graph should expose stable identifiers, provenance records, license terms, and machine-readable schema. Governance structures must also specify roles for curators, domain experts, data stewards, and model validators. Without such institutional arrangements, graph quality degrades over time, and predictions become detached from the evidence base. The governance challenge is not solved by a single tool; it requires enduring organizational commitments to maintenance, documentation, and audit.

The fairness and robustness concerns interact in important ways. A model that is robust in aggregate terms may still perform poorly for rare diseases because the number of training edges is small. Conversely, efforts to improve fairness by adding data for underserved conditions may reduce aggregate accuracy if the new data are noisy or poorly standardized. These trade-offs should be made explicit through multi-objective evaluation frameworks that report performance by disease category, population group, and evidence density. Such reporting encourages system builders to treat equity as a system-level property rather than a post hoc adjustment.

## **7. Deployment, Sustainability, and Policy Implications**

Deploying a knowledge graph-based repurposing system in a translational setting requires alignment with biomedical and regulatory expectations. Predictions from such systems are typically hypotheses rather than actionable clinical recommendations. They must be communicated to medicinal chemists, preclinical researchers, and clinical investigators in ways that convey mechanistic rationale, evidential strength, and uncertainty. Reviews of drug repurposing have repeatedly noted that computational predictions are only the beginning of a long validation process [14]. Therefore, deployment platforms should support workflow

integration, evidence capture, and collaborative review rather than functioning as isolated scoring engines.

Sustainability depends on the maintenance of both the graph and the model ecosystem. Ontologies such as the Gene Ontology [5] and the Human Phenotype Ontology [6] require continuous community curation; if these upstream resources stagnate, the downstream graph loses semantic resolution. Collaborative knowledge bases such as Wikidata [7] demonstrate the potential of community-driven maintenance, but biomedical resources need specialized expertise and careful conflict resolution. Funding models that only support initial construction without long-term curation create a familiar infrastructural failure in which systems become outdated and lose trust. Institutional sustainability should therefore be designed as part of the initial architecture rather than treated as an afterthought.

Policy implications extend from data sharing to clinical translation. Biomedical knowledge graphs often contain data derived from multiple jurisdictions, each with different intellectual property, privacy, and research ethics requirements. Aggregating these data into a unified graph can create regulatory ambiguity, especially when predictions cross national boundaries or involve patient-derived data. Aligning graph governance with FAIR principles [20] can improve interoperability, but it does not automatically resolve questions of consent, commercial confidentiality, or liability. Policymakers, research funders, and institutional review boards should develop guidance that recognizes knowledge graphs as dynamic research infrastructures with distributed responsibilities.

The broader societal impact of graph-based drug repurposing includes the possibility of redirecting research toward underrepresented diseases. Because graphs make hidden connections visible, they can help identify plausible treatments for conditions that lack commercial incentives. Achieving this potential requires deliberate investment in data coverage, curation for neglected diseases, and fairness-aware evaluation [19]. It also requires publishing negative results and failed predictions so that the graph does not become dominated by positive findings. A mature policy framework would treat such negative evidence as a valuable public good and support its structured integration into biomedical knowledge graphs.

The long-term value of graph-based repurposing depends on trust. Trust arises not only from predictive accuracy but also from transparency, auditability, and responsiveness to correction. Model documentation [19], provenance tracking, evidence attribution, and versioned releases all contribute to a system that can be interrogated by scientists and regulators. When predictions are presented with clear explanations and caveats, domain experts can assess whether a proposed relationship is mechanistically plausible and worth experimental follow-up. This sociotechnical integration is essential because the ultimate goal is not to replace human judgment but to extend it with structured, governed, and transparent knowledge.

## **8. Conclusion**

Knowledge graph-based modeling of compound–target–disease associations offers a powerful framework for drug repurposing, but its value depends on more than algorithmic sophistication. The graph must be designed as a governed infrastructure with robust identity resolution, ontology alignment, provenance tracking, and version control. Inference methods must balance interpretability, scalability, and robustness under incomplete knowledge. Fairness and policy considerations must be embedded in the system from the outset to ensure that repurposing opportunities are not disproportionately concentrated in well-studied disease

areas. Sustainable deployment requires long-term curation, institutional data stewardship, and regulatory alignment. When these sociotechnical dimensions are addressed together, knowledge graphs can become reliable engines for generating transparent and equitable therapeutic hypotheses.

## References

1. Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
2. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. von. (2019). STRING v11: Protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
3. Davis, A. P., Grondin, C. J., Johnson, R. J., Sciaky, D., Wieggers, J., Wieggers, T. C., & Mattingly, C. J. (2021). Comparative Toxicogenomics Database (CTD): Update 2021. *Nucleic Acids Research*, 49(D1), D1138–D1143. <https://doi.org/10.1093/nar/gkaa891>
4. Piñero, J., Ramírez-Anguaita, J. M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., & Furlong, L. I. (2020). The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Research*, 48(D1), D845–D855. <https://doi.org/10.1093/nar/gkz1021>
5. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: Tool for the unification of biology. *Nature Genetics*, 25(1), 25–29. <https://doi.org/10.1038/75556>
6. Köhler, S., Gargano, M., Matentzoglou, N., Carmody, L. C., Lewis-Smith, D., Vasilevsky, N. A., Danis, D., Balagura, G., Baynam, G., Brower, A. M., Callahan, T. J., Chute, C. G., Est, J. L., Galer, P. D., Ganesan, S., Griese, M., Haimel, M., Pazmandi, J., Hanauer, M., ... Robinson, P. N. (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49(D1), D1207–D1217. <https://doi.org/10.1093/nar/gkaa1043>
7. Vrandečić, D., & Krötzsch, M. (2014). Wikidata: A free collaborative knowledgebase. *Communications of the ACM*, 57(10), 78–85. <https://doi.org/10.1145/2629489>
8. Hogan, A., Blomqvist, E., Cochez, M., d’Amato, C., de Melo, G., Gutierrez, C., Kirrane, S., Gayo, J. E. L., Navigli, R., Neumaier, S., Ngonga Ngomo, A.-C., Polleres, A., Rashid, S. M., Rula, A., Schmelzeisen, L., Sequeda, J., Staab, S., & Zimmermann, A. (2021). Knowledge graphs. *ACM Computing Surveys*, 54(4), 1–37. <https://doi.org/10.1145/3447772>
9. Wang, Q., Mao, Z., Wang, B., & Guo, L. (2017). Knowledge graph embedding: A survey of approaches and methods. *IEEE Transactions on Knowledge and Data Engineering*, 29(12), 2724–2743. <https://doi.org/10.1109/TKDE.2017.2754499>

10. Nickel, M., Murphy, K., Tresp, V., & Gabrilovich, E. (2016). A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1), 11–33. <https://doi.org/10.1109/JPROC.2015.2483592>
11. Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., & Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. *Advances in Neural Information Processing Systems*, 26, 2787–2795.
12. Teru, K., Denis, E., & Hamilton, W. L. (2020). Inductive relation prediction by subgraph reasoning. *Proceedings of the 37th International Conference on Machine Learning*, 9448–9457.
13. Himmelstein, D. S., Lizee, A., Hessler, C., Brueggeman, L., Chen, S. L., Hadley, D., Green, A., Khankhanian, P., & Baranzini, S. E. (2017). Systematic integration of biomedical knowledge prioritizes drugs for repurposing. *eLife*, 6, e26726. <https://doi.org/10.7554/eLife.26726>
14. Pushpakom, S., Iorio, F., Eyers, P. A., Escott, K. J., Hopper, S., Wells, A., Doig, A., Williams, T., Latimer, J., McNamee, C., Norris, A., Sanseau, P., Cavalla, D., & Pirmohamed, M. (2019). Drug repurposing: Progress, challenges and recommendations. *Nature Reviews Drug Discovery*, 18(1), 41–58. <https://doi.org/10.1038/nrd.2018.168>
15. Lotfi Shahreza, M., Ghadiri, N., Mousavi, S. R., Varshosaz, J., & Green, J. R. (2018). A review of network-based approaches to drug repositioning. *Briefings in Bioinformatics*, 19(5), 878–892. <https://doi.org/10.1093/bib/bbx017>
16. Wang, Yinyin, et al. "HerbSyner\_Finder: a network community-based model for identifying synergistic combinations from herbal medicines and complex systems." *Targetome 2.2* (2026).
17. Zitnik, M., Agrawal, M., & Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics*, 34(13), i457–i466. <https://doi.org/10.1093/bioinformatics/bty294>
18. Hamilton, W. L., Ying, R., & Leskovec, J. (2017). Inductive representation learning on large graphs. *Advances in Neural Information Processing Systems*, 30, 1024–1034.
19. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. <https://doi.org/10.1145/3287560.3287596>
20. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>