

Insights from Knowledge Graphs : Introducing a new formalism

1. Insights
What are insights?

2. Commonly used methods to gain insights
Reasoning, Smart Queries ...

3. Knowledge Graphs to Knowledge Networks: Formalism
Formalism to convert KGs to Unipartite Knowledge Networks
A knowledge graph is represented by a set of triples, which consist of subjects, predicates, and objects.
 $KG = \{S, P, O\}$
A unipartite network KG_u consists of set of nodes V and edges E
 $N_u = (V, E)$
For a Knowledge Graph $KG = \{S, P, O\}$
 $2P, P_u$
 W_{P_u}
 $N_{P_u} = (V_u, E_{P_u})$
 $N_u = N_v + N_{P_u}$
Future Work
Explore and create formalisms for knowledge graphs

4. Insights from Knowledge Networks: Visual Analytics
What is seen vs What is encoded

5. Insights from Knowledge Networks: Network Science Approach
Network Metrics

Metric	Earth	Mars	Moon	Venus
Density	0.04	0.27	0.00	0.76
Maximum Network Diameter	3	5	5	3

LIVE SESSION | ABSTRACT | REFERENCES | CONTACT AUTHOR | PRINT | GET POSTER

Anirudh Prabhu*(prabha2@rpi.edu,@anirudh_14) and Peter Fox*
(pfox@cs.rpi.edu,@taswegian)

*Tetherless World Constellation, Rensselaer Polytechnic Institute, Troy, NY, USA.



PRESENTED AT:



1. INSIGHTS

What are insights?



How to gain insights?

How do we gain insights?

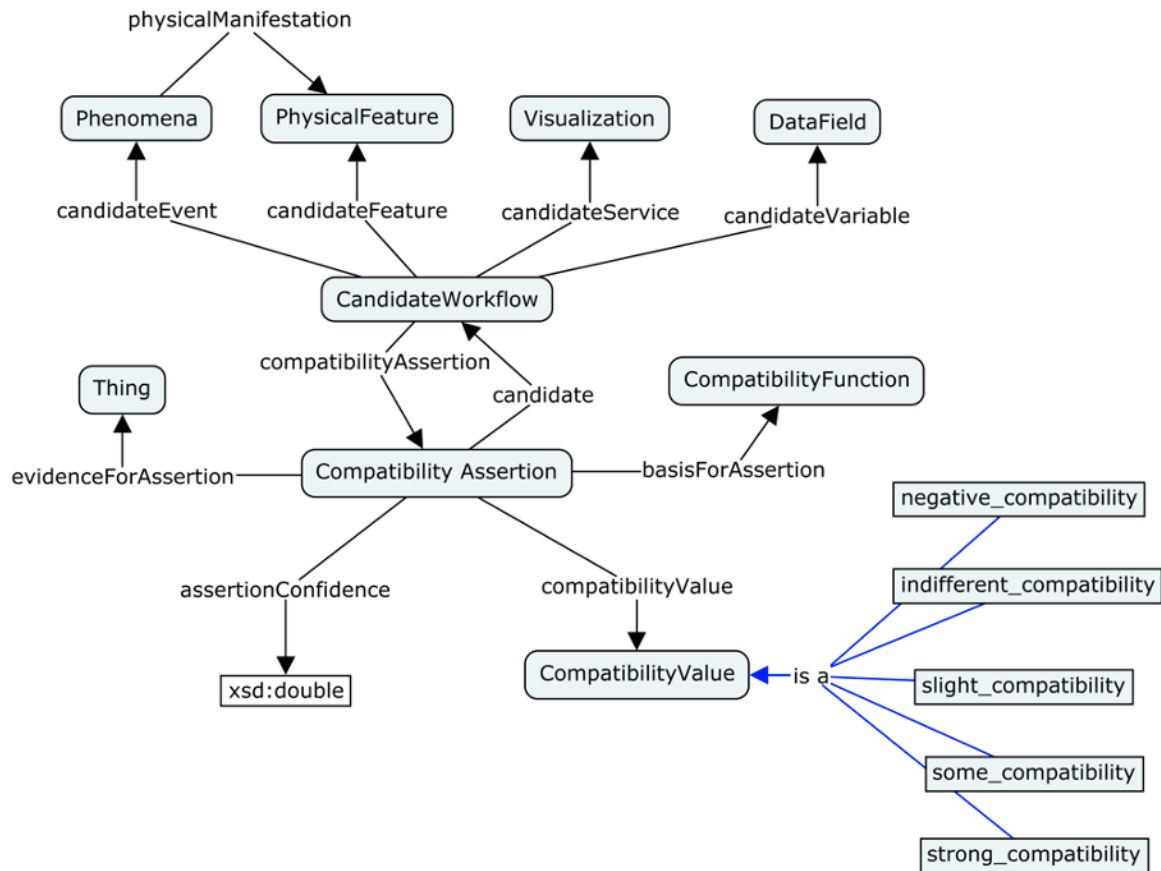
Reasoners

Visual Analytics

Network Science Approach

2. COMMONLY USED METHODS TO GAIN INSIGHTS

Reasoning, Smart Queries



Conceptual Map of Ontology

Rules Engine – Apache Jena Example

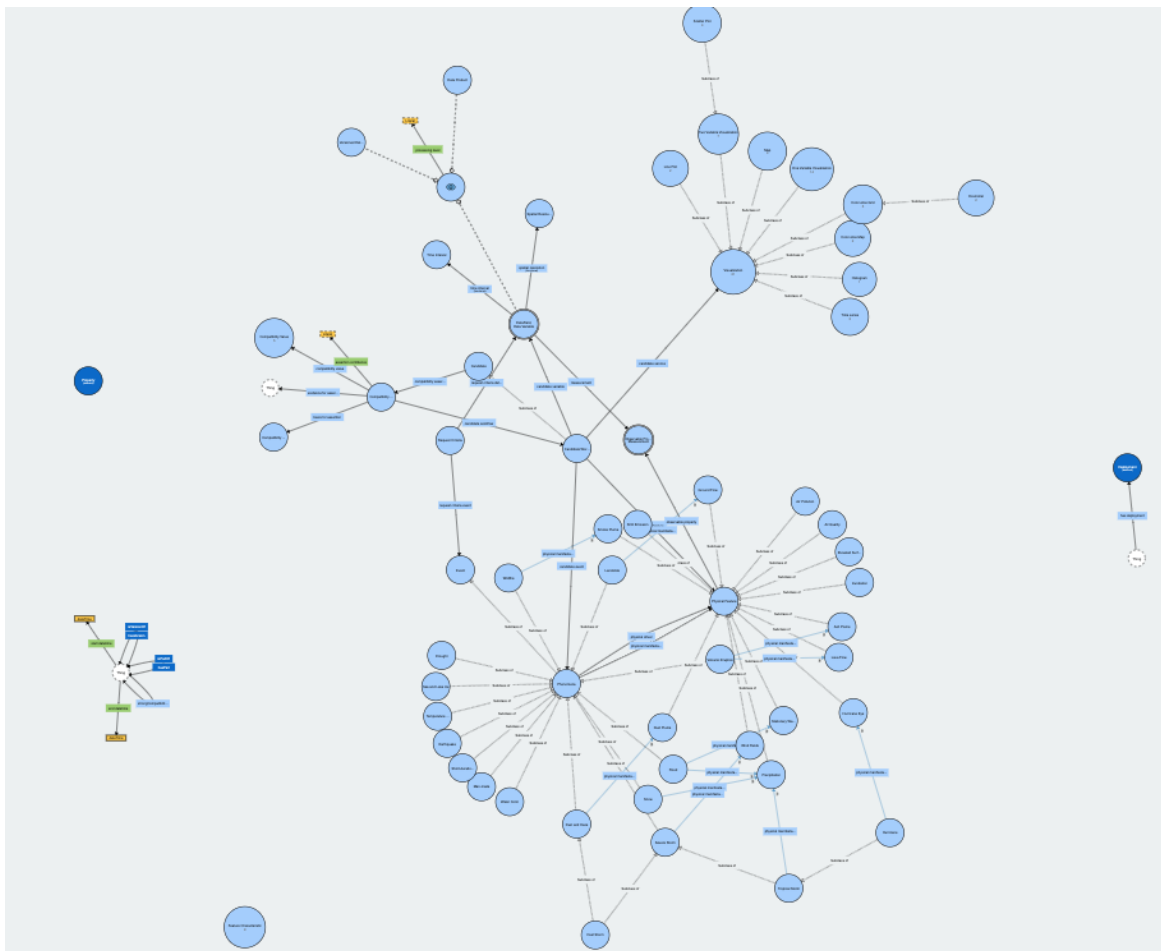
'Half hourly' time interval is best for Hurricanes and Tropical Storms.

```
[hurricane-half-hourly:
  (?candidate dd:candidateEvent ?event),
  (?event rdf:type dd:Hurricane),
  (?candidate dd:candidateVariable ?variable),
  (?variable dd:timeInterval ?timeInterval),
  equal(?timeInterval, <http://darkdata.tw.rpi.edu/data/time-interval/half-hourly>),
  makeSkolem(?assertion, dd:Hurricane, ?timeInterval)
->
  (?candidate dd:compatibilityAssertion ?assertion),
  (?assertion rdf:type dd:CompatibilityAssertion),
  (?assertion dd:compatibilityValue dd:strong_compatibility),
  (?assertion dd:assertionConfidence "0.5"^^xsd:double),
  (?assertion dd:basisForAssertion <urn:rule/time_interval/hurricane-half-hourly>)
]
```

Antecedent :
Containing
information about
Phenomena and
Temporal
Resolution.

Subsequent :
Containing the
compatibility
assertion
information.

Example of rules used in Ontology



Ontology visualized using VOWL (<http://www.visualdataweb.de/webvowl/>) (<http://www.visualdataweb.de/webvowl/>)

3. KNOWLEDGE GRAPHS TO KNOWLEDGE NETWORKS : FORMALISM

Formalism to convert KGs to Unipartite Knowledge Networks

A knowledge graph is represented by a set of triples, which consist of subjects, predicates and objects.

$$KG = (S, P, O)$$

A unipartite network(N_u) consists of vertices(V) and edges(E)

$$N_u = (V, E)$$

For a Knowledge Graph $KG = (S_i, P_j, O_k)$

$$\exists! P,$$

$$\forall O_1^k,$$

$$N'_u = \left(V_i, E_{\frac{n(n-1)}{2}} \right),$$

$$N_u = N_u + N'_u$$

Future Work

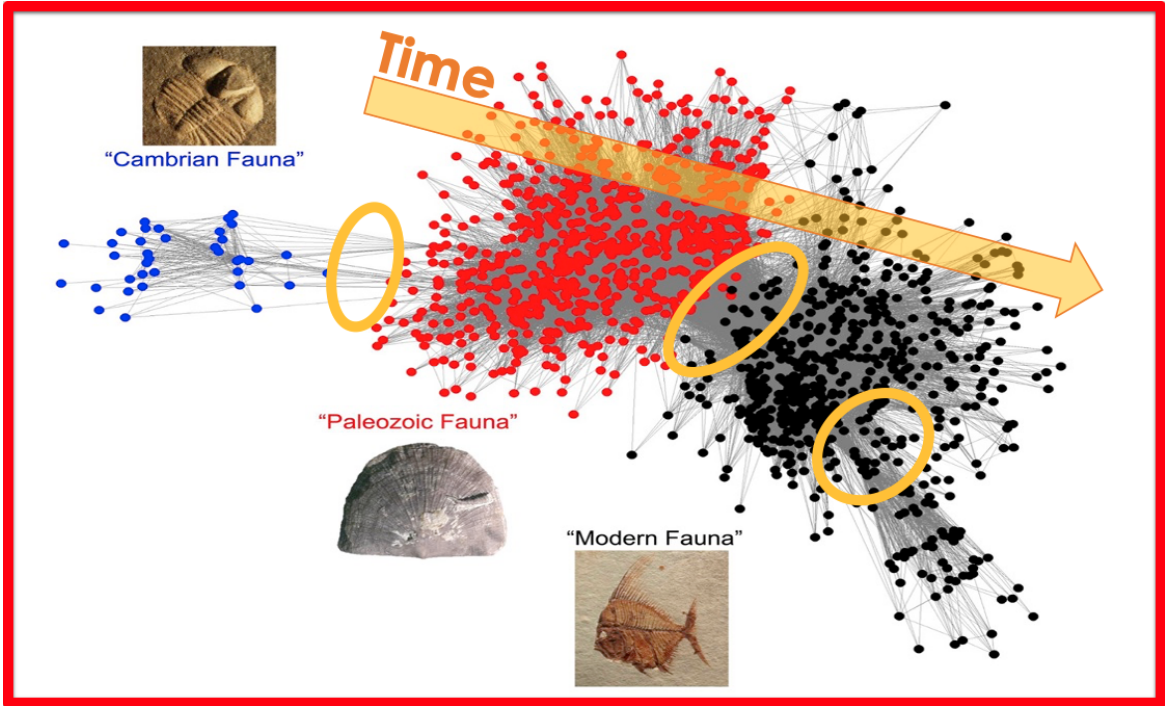
Explore and create formalisms for knowledge graphs to bipartite and other k-partite networks.

4. INSIGHTS FROM KNOWLEDGE NETWORKS: VISUAL ANALYTICS

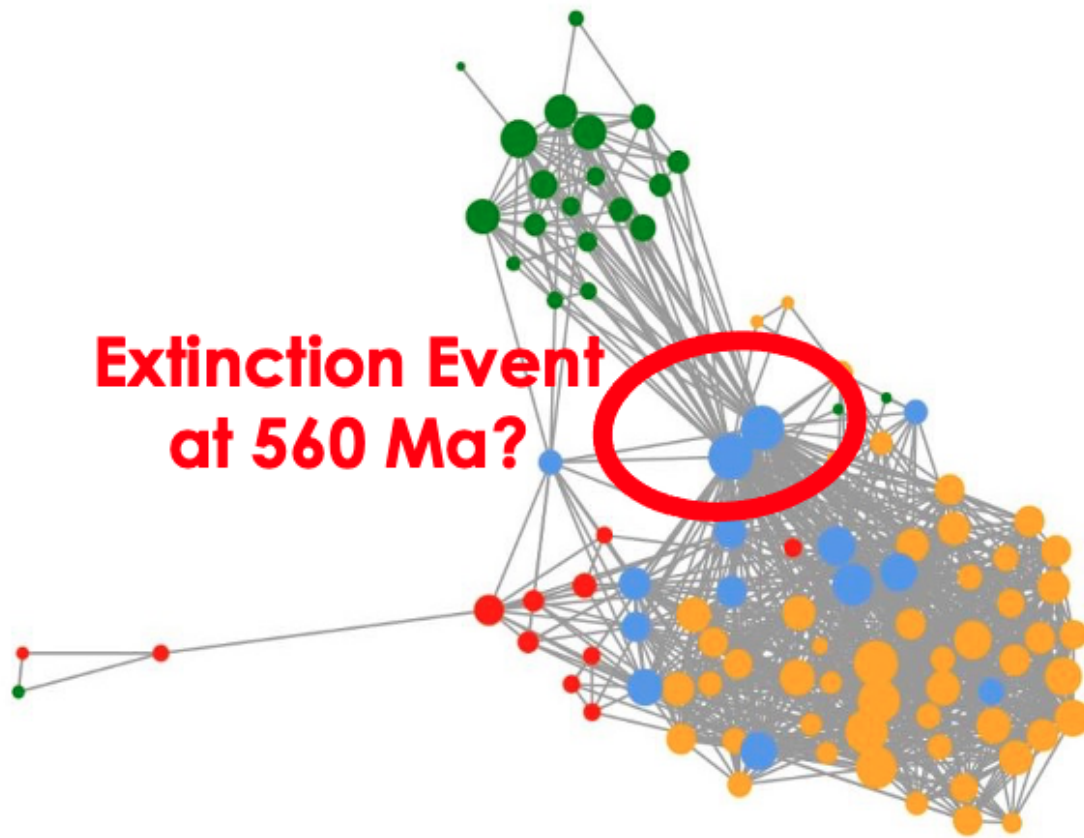
What is seen vs What is encoded

Encoded	Seen/Inferred/Calculated
Nodes	Patterns in the Network Geometry
Edges	Sub-Communities formed in the Network
Layout (Mostly Force Directed)	Important Hubs in the Network
Additional Parameters for Nodes (Optional)	Additional metrics that explain the complexity of the environment (<u>assortativity</u> , betweenness, centrality etc.)

In the visual layout



Animal Family Fossil Network (from 542 Ma - present)



Ediacaran Fossil Network (From 635 Ma to 541 Ma)




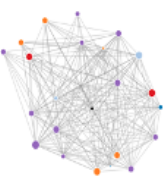
Evolving Networks

[VIDEO] <https://www.youtube.com/embed/x1P3FZhWALw?rel=0&fs=1&modestbranding=1&rel=0&showinfo=0>

Network of Copper Minerals evolving through time. (3.7 Ga to 2.4 Ga)

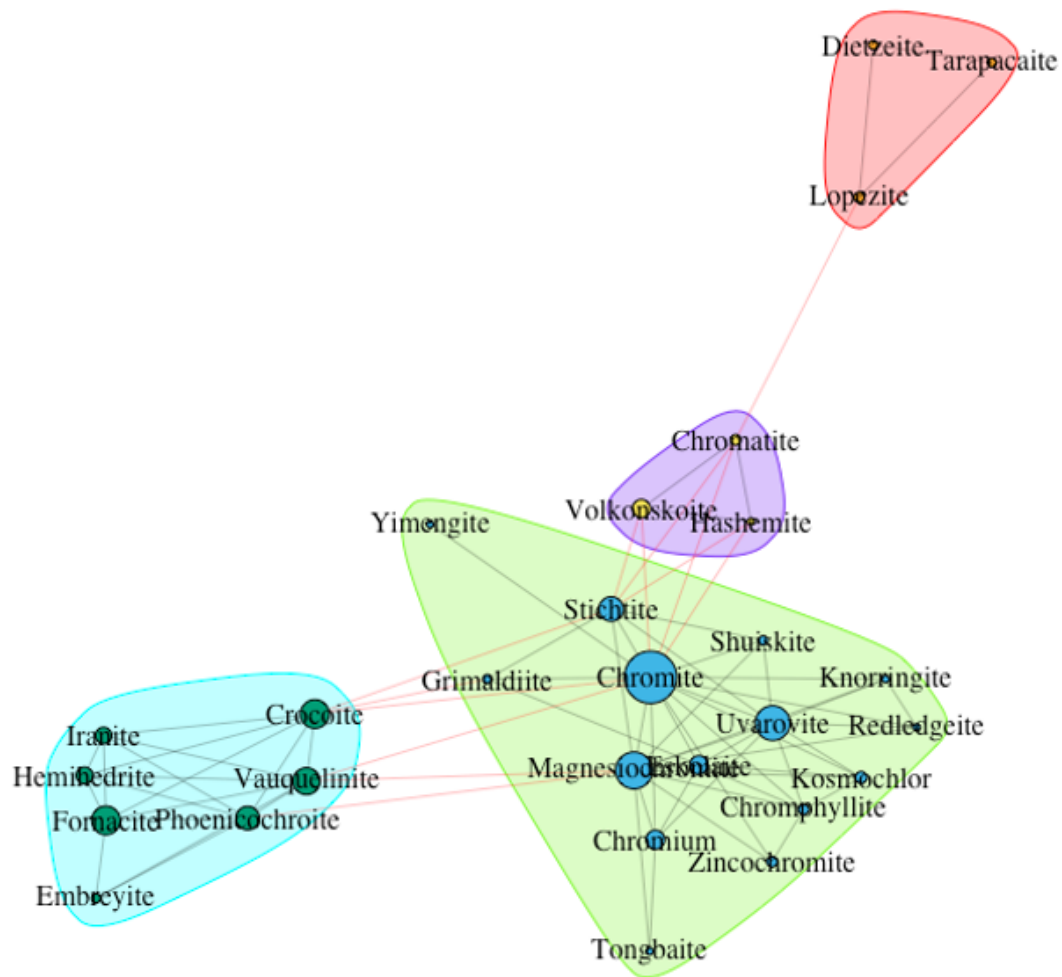
5. INSIGHTS FROM KNOWLEDGE NETWORKS: NETWORK SCIENCE APPROACH

Network Metrics

				
Metric	Earth	Mars	Moon	Vesta
Density	0.64	0.27	0.69	0.78
Maximum Network Diameter	2	3	3	2
Mean Network Diameter	1.36	1.69	1.26	1.22
Degree Central.	0.34	0.62	0.23	0.22
<u>Betweenness</u> Centralization	0.02	0.09	0.03	0.02

Comparing Global Metrics in mineral networks of planetary bodies

Community Detection Algorithms



Walktrap algorithm applied to Chromium Mineral Network.

- Groups correspond to Paragenetic Mode.
- Paragenetic Mode : Formation Conditions.
- How and when the Minerals were formed.

ABSTRACT

The usage of the term Knowledge Graph (KG) has gained significant popularity since 2012, when Google introduced its own knowledge graph, and how they used it to enhance their searches and question answering systems. While various definitions and interpretations for knowledge graphs have been presented, what remains consistent is that knowledge graphs are commonly used with reasoners to make inferences about data, based on assertions and axioms written by human experts. But knowledge graphs, which store complex, multi-dimensional data contain hidden patterns and trends that cannot be explored simply using reasoners. In such a case it becomes necessary to extract parts of the knowledge graph (focusing on the instances related to one property at a time) and analyze them individually in order to conduct a focused but tractable exploration of the domain. In this presentation, we present one way to gain insights from knowledge graphs, using network science. To achieve this goal, we have formalised the partitioning of knowledge graphs to unipartite knowledge networks, and present various ways to explore and analyse such knowledge networks to form scientific hypotheses, gain scientific insights and make discoveries.

REFERENCES

- <https://towardsdatascience.com/knowledge-graphs-and-machine-learning-3939b504c7bc>
- M. Kroetsch and G. Weikum. Journal of Web Semantics: Special Issue on Knowledge Graphs. <http://www.websemanticsjournal.org/index.php/ps/announcement/view/19> [August, 2016].
- Pujara, J., Miao, H., Getoor, L., & Cohen, W. (2013). Knowledge graph identification. In Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-642-41335-3_34
- Färber, M., Bartscherer, F., Menne, C., & Rettinger, A. (2018). Linked data quality of DBpedia, Freebase, OpenCyc, Wikidata, and YAGO. Semantic Web. <https://doi.org/10.3233/SW-170275>
- Paulheim, H. (2017). Knowledge graph refinement: A survey of approaches and evaluation methods. Semantic Web. <https://doi.org/10.3233/SW-160218>
- Moore, E. K., Hao, J., Prabhu, A., Zhong, H., Jelen, B. I., Meyer, M., ... & Falkowski, P. G. (2018). Geological and Chemical Factors that Impacted the Biological Utilization of Cobalt in the Archean Eon. Journal of Geophysical Research: Biogeosciences.
- Morrison SM, Liu C, Eleish A, Prabhu A, Li C, Ralph J, Downs RT, Golden JJ, Fox P, Hummer DR, Meyer MB, and Hazen RM (2017) Network analysis of mineralogical systems. American Mineralogist 102
- Muscente, A. D., Prabhu, A., Zhong, H., Eleish, A., Meyer, M. B., Fox, P., ... & Knoll, A. H. (2018). Quantifying ecological impacts of mass extinctions with network analysis of fossil communities. Proceedings of the National Academy of Sciences, 115(20), 5217-5222.
- Ehrlinger, L., & Wöß, W. (2016). Towards a Definition of Knowledge Graphs. SEMANTiCS (Posters, Demos, SuCCESS), 48, 1-4.
- Blumauer, A. (2014). From Taxonomies over Ontologies to Knowledge Graphs. Semantic Web Company.