Importance-driven focus+context for inspiration analysis in citation graphs

Marius Irgens

University of Bergen, Norway

Abstract

This paper introduces importance-driven citation graph rendering as a novel technique for focus+context display of inspiration. It uses author-to-author citations as an example: How inspired is a source-author by the other authors, and how inspired are the other authors by that source-author. However, the concept is not exclusive to authors, and could be applied to any set of citation givers and takers; science groups or scientific journals for instance.

CCS Concepts

• ***Human-centered computing*** $\rightarrow$ ***Graph drawings;***

1. Introduction

A citation graph is a directed graph that describes the distribution of citations within a set of citation givers and receivers. The citation givers and receivers can be papers, authors, science groups, or even scientific journals.
Citation graphs are often big and dense, with many overlapping nodes and edges. To be able to analyse such a graph, interactive visualization is crucial. Even in simple usecases, the visualization mantra "Overview first, zoom and filter, then details-on-demand" become important as nodes and edges start to overlap. If the goal is to visually analyse single elements in such crowded graphs, interactivity becomes even more important. But if the visualization *was* interactive, what kind of interactive functionality would provide most useful? It depends on exactly what is to be analysed.
As mentioned in the abstract, this paper will focus on analysis of author *inspiration*. To do this, one needs tools that can isolate the inspiration of specific authors, as well as communicate inspirational degree within that isolated area. To isolate, I use concepts from the focus+context paradigm. To communicate inspirational degree, I use concepts from the importance-driven rendering paradigm.

1.1. What is inspiration?

There are two main categories of inspiration that should be visualized: an authors inspiration *from* other authors, and an authors inspiration *to* other authors. A third category of reciprocal inspiration emerge when authors belong to both. How can one define author-to-author inspiration mathematically? I propose the following:
1. An author is inspired by another author if his total number of citations towards that author exeeds some user-defined minimum threshold. I will use a threshold of 8 citations for my examples, but this really depends on the data, and should ultimately be user-definable.
2. An author is indirectly inspired by another author if he is inspired by someone who is inspired by that author. This relation is recursive, and the indirect inspiration distance from one author to another equals the least number of inspirational jumps needed to get there (the shortest path).
3. The inspirational degree is further defined by the number of citations in an inspiration connection.
When a source-author is defined, who he is inspired by can be visualized as a tree. Which authors are inspired by the source-author can likewise be visualized as a tree, by using the same logic in the opposite direction. Something very interesting happens when visualizing both at the same time that requires special treatment - cyclic graphs emerge (figure 1).

1.2. why is inspiration useful?

Visualization of inspiration can be used to:

- Evaluate authors contribution to a community.
- Investigate author group dynamics.
- Analyse substructures of inspiration graphs.

2. Related Work

2.1. Focus+context

Applying focus+context concepts to a citation graph can help a user understand the inspiration of a selected source-author, compared to the total amount of inspiration in the author set. Each author contains a subgraph consisting of an incoming inspiration tree, an outgoing inspiration tree, and a number of cyclic graphs of

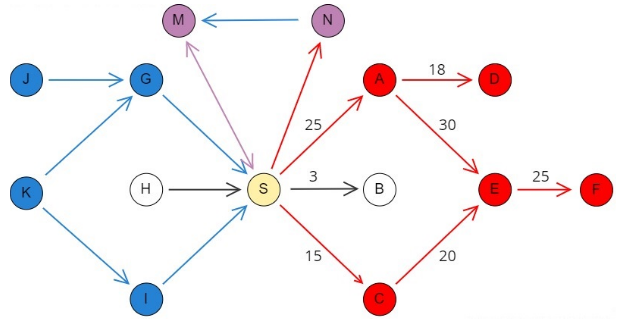

Figure 1: *Author S is the source. Author B is only cited 3 times and therefore not counted as inspiring. Authors A and C are directly inspiring. Authors D, E and F are indirectly inspiring. The red nodes represent authors that author S is inspired by. The blue nodes represent authors inspired by author S. The purple nodes represent authors in a reciprocal inspirational relationship: Author S is directly inspired by author M, and author M is directly inspired by author S back. Author S is directly inspired by author N, but author N is only indirectly inspired by author S back. The purple edge represents equidistant reciprocal inspiration.*

reciprocal inspiration. This subgraph can be defined as a selectable focus object. All authors and citation edges *not* included in it is considered context. Separate rendering techniques can be used on focus and context to heighten the information bandwidth. In "Generalizing Focus+Context Visualization", Hauser et. al. mentions color, opacity and style as possible graphical resources to differentiate focus from context [Hau06].

2.2. Importance-driven rendering

Importance-driven rendering concepts can be applied to give the user better control of the focus reach and the render characteristics within it to heighten the information bandwidth further. In "Degree-of-interest trees: a component of an attention-reactive user interface" [CN02] by Card et. al, the intrinsic importance of a node (its "degree-of-interest") is defined as its distance from a focus node. The distance is then defined as "the number of nodes that must be traversed by following parent and child links from the node of interest until reaching the subject node." This value can then be used to drive the characteristics of the render, for instance its color, opacity, style, or geometric transformation.

2.3. Dynamic graph layout

In "A survey of two-dimensional graph layout techniques for information visualisation" [GFV13] by Gibson et. al, several graph layout considerations is discussed: "how the graph is drawn has a significant impact on how the graph is understood". The most important concepts are perceived relationship from proximity, perceived relationship from similarity, and uniform edge lengths.
In "Animated Exploration of Graphs with Radial Layout" [YFDH01] by Yee et. al, the graph layout changes dynamically, depending on which node is is selected as the source-node. Nodes are arranged radially on circles with radius equal to their distance from the source-node, clearly displaying parent-child hierarchies. Animation helps the user understand transitions between states.
In "H3: Laying Out Large Directed Graphs in 3D Hyperbolic Space" [Mun97] Munzner et. al use a similar approach to lay out nodes radially in 3D instead.

3. Method

3.1. Adjustable degree-of-inspiration

When a source-author is selected, all other authors can get assigned a degree-of-inspiration value based on their intrinsic importance to the source-author. First, check if the target-author is part of the focus by checking if it can be reached at all. If there is no path from the source-author to the target-author it has zero importance, and there is no need for further calculation - it is part of the context. Categorize the target-author as someone who *inspires* the source-author, *is inspired* by the source-author, or *both*. Then, based on the target-authors path distance from the source-author, and the number of citations in each step between them, calculate the degree-of-inspiration value. I came up with the following equation:

*insp(x) = inspMax(parent(x)) - (distance(x)/distanceK) - (distance(x)/numCit(x)*citK)*

insp(x) is the resulting degree-of-inspiration of a target-author. *Parent(x)* are all authors in focus with distance = (distance(x) - 1) that also lead to the target-author. *inspMax(parent(x))* is the highest degree-of-inspiration value of any parent of the target-author. *Distance(x)* is the path distance from the source-author to the target-author. *DistanceK* can be adjusted by the user. *NumCit(x)* is the total number of citations to the target-author from his parents. *CitK* can be adjusted by the user. Figure 2 shows an example calculation.
This results in a curved falloff. It is possible to make the falloff linear by replacing *distance(x)* with a constant number. Increasing the *distanceK* value makes distance less influential. Increasing the *citK* value makes the number of citations less influential. Increase or lower both to affect the reach of inspiration.
K-value adjustment should be interactive, with lowest possible computation cost to keep the framerate high. There is no need to perform the calculation for every author in the whole set. It is enough to perform it for every author in the focus subset. Parents should also be calculated before children, to keep from having to do unneccessary, recursive calculations (in the spirit of memoization). I suggest the following setup:
1. When a source-author is selected, make one list of all authors that can be reached from the source-author by following a shortest path of outgoing citations (reachable nodes), and a list of the edges inbetween (reachable edges). Make one list of all authors that can reach the source-author by following a shortest path of incoming citations (trailing nodes), and a list of the edges inbetween (trailing edges). Make one list of authors that belong to both (cyclic nodes), and a list of the edges inbetween (cyclic edges). Each node-list element should contain distance, a list of parents, and total number

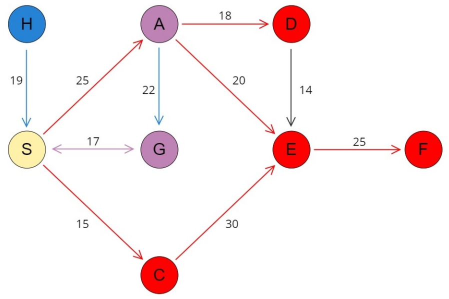

Figure 2: *Author S has degree-of-inspiration (DOI) = 1. The user has chosen distanceK = 10 and citK = 1.*
Author A has distance = 1 and numCit = 25. His resulting DOI = 1 - 1/10 - 1/25 = 0.86. Author C gets DOI = 0.83.
Author E has distance = 2 and numCit = 50. His resulting DOI = 0.86 - 2/10 - 2/50 = 0.62. Author D gets DOI = 0.55.
The edge DE is disregarded since author E has distance = 2, and this edge has distance = 3.
Author G has distance = 1 and numCit = 39. His resulting DOI = 1 - 1/10 - 1/39 = 0.87.
Author H ha distance = 1 and numCit = 19. His resulting DOI = 1 - 1/10 - 1/19 = 0.85.

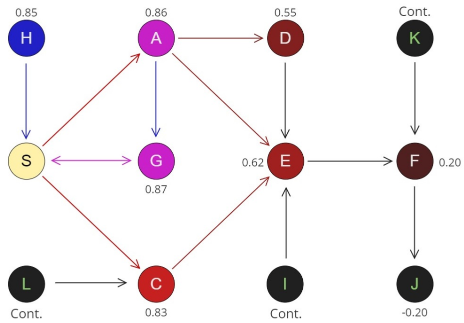

Figure 3: *A transfer-function takes inspiration-type and DOI value as argument and returns a color value. Authors I, K and L is not part of author S' focus, and is assigned a context material. Author J's DOI-value falls below threshold = 0, and is also assigned a context material.*

of citations from all parents. Cyclic node-list elements is made by combining the information from both the referenced reachable node-list element and the trailing node-list element: concatenate the lists of parents, add together the number of citations, and use the lowest distance value of the two.
2. Sort these lists by distance, ascending.
3.When the user changes the K-values, iterate through these lists in ascending fashion and perform the degree-of-inspiration calculations. Start with reachable and trailing nodes, then cyclic nodes. This will make sure all parents are calculated before children.

3.2. Inspiration-driven transfer-function

A transfer function takes the authors inspiration-type and degree-of-inspiration as arguments and returns a render setting. The inspiration-type says which list he belongs to: reachable nodes, trailing nodes or cyclic nodes - these should be rendered differently. The degree-of-inspiration can be used to interpolate between material traits (like color) or other rendering characteristics.
The edges inbetween can use the degree-of-inspiration value of the node they "belong to" (meaning the author that uses its citation value) for their transfer function. Their directionality should also be expressed clearly: reachable edges, trailing edges and cyclic edges should be rendered differently.
All nodes and edges that does not belong to the focus, and therefore is part of the context, can use a context render setting. In addition, any node or edge within the focus that gets a degree-of-inspiration value below some chosen threshold can be assigned that same context render setting. In this way, the user can adjust the reach of the focus as well as its render characteristics by changing the K-values (figure 3).

Cycles can get created or broken by updating the K-values. The transfer function needs to perform an additional check for every cyclic node, to see if it really is part of a cycle with respect to the current K-values. This can be solved with a recursive backtracking function in both directions, over parents with positive DOI values, that returns true if the source has been reached, or false otherwise (both directions must reach the source). This might sound computational heavy, but unless the focus graph is very heavily populated with cyclic nodes it should not affect performance much.

3.3. Distance-driven force layout

When the graph is constructed it is assigned a default layout. A static layout can make it hard to trace inspiration paths in a focus. A dynamic layout that can update with respect to the selected source-author, ordered in a parent-child hierarchy, is helpful. A popular dynamic layout concept is force-directed layouts - nodes are moved by simulating forces on them. A good force simulation supports both animated movement and interactive adjustments of force attributes. One choice could be the popular force simulation API d3-force-3d [d3f]. It is highly optimized and supports definition of custom forces. D3-force-3d is a 3D extension of the 2D force simulation d3-force, which is part of the popular API D3.js. This is how one could use it to obtain a distance-driven force layout:
The default graph layout is made using forces that are provided with the API - center force, link length force and collision force. The center force makes sure every node gravitates towards a defined center. The link length force uses spring-force simulation to try to maintain equal edge lengths between connected nodes. The collision force preserves a minimum distance between all nodes so they do not overlap (figure 8).
In addition to these forces, customize a radial force that rearranges the nodes to make parent-child relationships clearer, by using their distance from the source-node as an argument (figure 4). D3-node-3d supports custom force definition that can be used in addition to the default forces, to create complex, dynamic layouts.

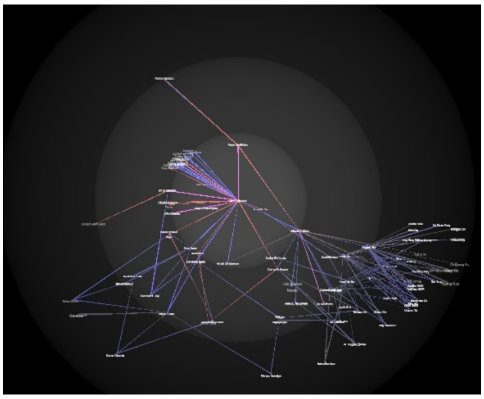

Figure 4: *A radial force dynamically rearranges the nodes to easen path-tracing: Move the source-author to the center. Every other author gravitates towards or away from the center, such that its radial length from the center equals its distance * link length.*

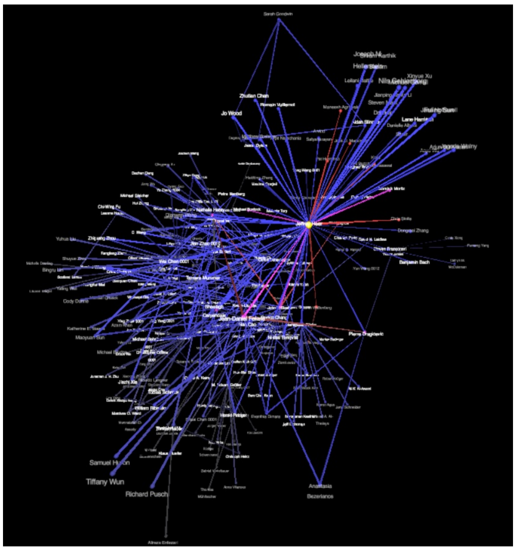

Figure 5: *The inspiration graph of Jeffrey Heer: clearly a highly inspirational individual. Not symmetrically inspired back.*

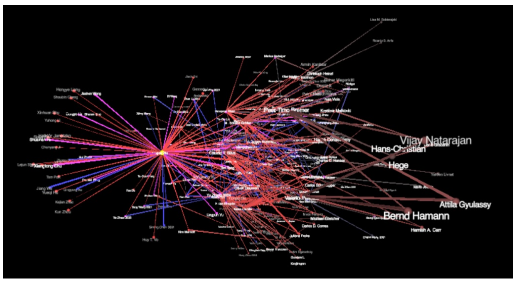

Figure 6: *The inspiration graph of Yingcai Wu: almost the opposite of Jeffrey Heer. Very inspired by the visualization community, with a high number of direct reciprocal inspiration as well.*

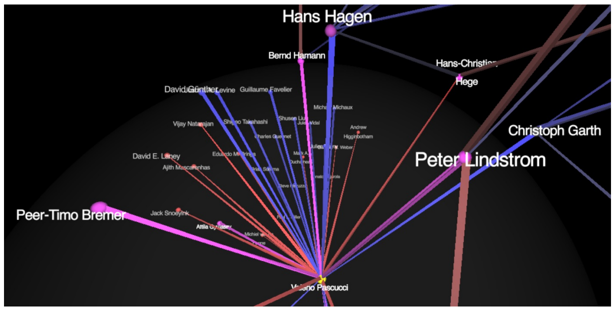

Figure 7: *Valerio Pascucci (bottom) is directly inspired by Hans-Christian Hege (top right), but Hans-Christian Hege is only indirectly inspired by Valerio Pascucci through Hans Hagen (top). Also, Hans Hagen is directly inspired by Valerio Pascucci, but Valerio Pascucci is only indirectly inspired by Hans Hagen through Hans-Christian Hege. This circle of inspiration is quite fascinating to me: they are all inspired by eachother, but at different qualities.*

4. Conclusions

Importance driven focus+context for inspiration analysis can be implemented as part of a citation graph application to evaluate individual authors, as shown in figure 5 and 6. Relatively complex group dynamics can also be revealed, as shown in figure 7. Substructures found in inspiration graphs can be compared to the substructures laid out by Yong et. al. in "Lattices in citation networks: An investigation into the structure of citation graphs" [YR01] for structure analysis (figure 12 and 13).

References

[CN02] Card S. K., Nation D.: Degree-of-interest trees: A component of an attention-reactive user interface. In *Proceedings of the Working Conference on Advanced Visual Interfaces* (2002), pp. 231–245. doi:doi.org/10.1145/1556262.1556300. 2

[d3f] d3-force-3d: force simulation API (github repository with documentation). https://github.com/vasturiano/d3-force-3d. Accessed: 2022-11-11. 3

[GFV13] Gibson H., Faith J., Vickers P.: A survey of two-dimensional graph layout techniques for information visualisation. *Information visualization 12*, 3-4 (2013), 324–357. doi:doi.org/10.1177/1473871612455749. 2

[Hau06] Hauser H.: Generalizing focus+ context visualization. In *Scientific visualization: The visual extraction of knowledge from data*. Springer, 2006, pp. 305–327. doi:doi.org/10.1007/3-540-30790-7_18. 2

[Mun97] Munzner T.: H3: Laying out large directed graphs in 3d hyperbolic space. In *Proceedings of VIZ'97: Visualization Conference, Information Visualization Symposium and Parallel Rendering Symposium* (1997), IEEE, pp. 2–10. doi:10.1109/INFVIS.1997.636718. 2

[YFDH01] Yee K.-P., Fisher D., Dhamija R., Hearst M.: Animated exploration of graphs with radial layout. In *Proc. IEEE InfoVis 2001* (2001), pp. 43–50. doi:doi.org/10.1109/INFVIS.2001.963279. 2

[YR01] Yong F., Rousseau R.: Lattices in citation networks: An investigation into the structure of citation graphs. *Scientometrics 50*, 2 (2001), 273–287. doi:https://doi.org/10.1023/A:1010573723540. 4, 5

Appendix

For my evaluation I used javascript ES6. The dataset is derived from the visualization publications dataset, and contains author-to-author citations with a minimum citation threshold of 8.

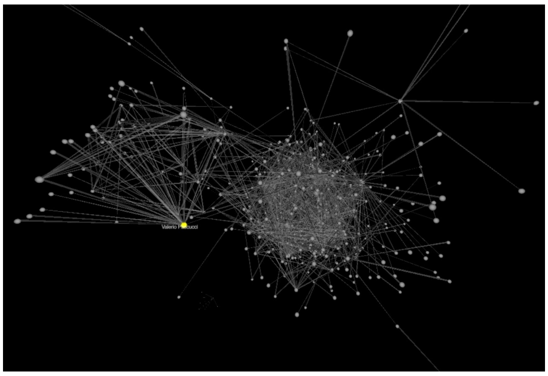

Figure 8: *The default layout is arranged by simulating forces on the nodes: center force, link length force and collision force.*

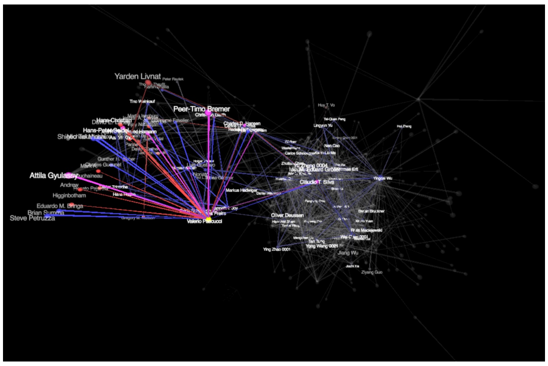

Figure 9: *Selecting an author constructs an opaque, colored focus with name labels. The context is rendered semi-transparently.*

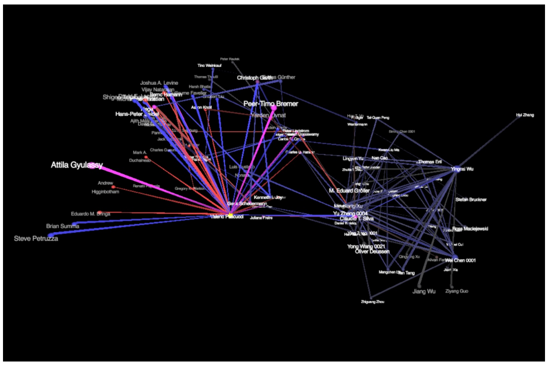

Figure 10: *By adjusting the opacity of the context, the user can choose when and how to view it - here with opacity = 0.*

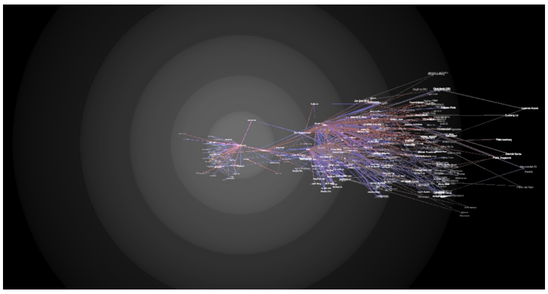

Figure 11: *Gaining the K-values can reveal a quite complex graph. compare this figure (with distanceK = 50 and citK = 4) to figure 4 (with distanceK = 25 and citK = 1).*

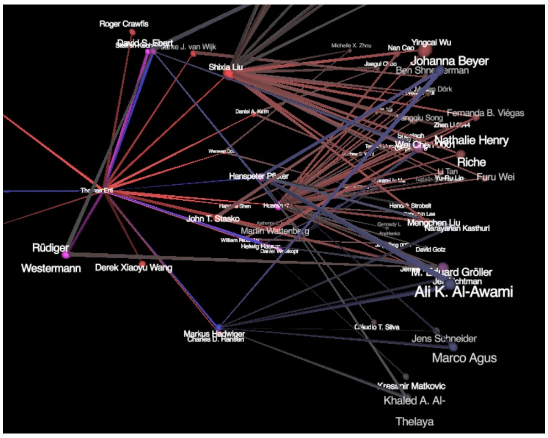

Figure 12: *The diamond shape is considered one of the most important substructures in citation graphs [YR01]. Thomas Ertls inspiration graph contains an incredible amount of them.*

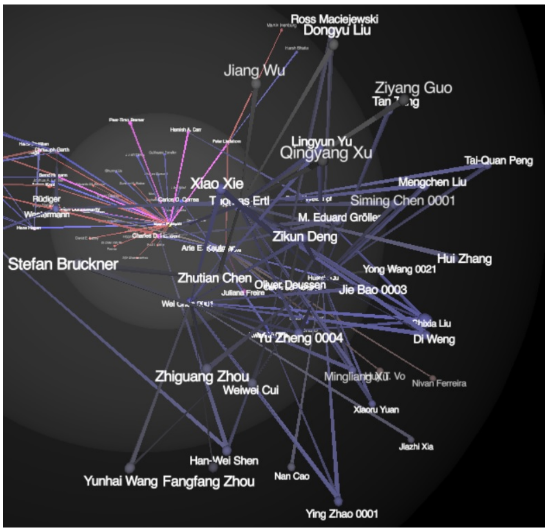

Figure 13: *Valerio Pascuccis inspiration graph contains this interesting looking crystal/star structure - itself consisting of multiple simpler substructures covered by Yong et. al. in [YR01]*