Propagation and aggregation of opinions in Structural Graphs

Carles Sierra, Pablo Noriega, Nardine Osman, Jordi Sabater,

December 15, 2009
Problems with current publication paradigm

- very competitive system for gaining reputation
  - researcher opposes early sharing due to fear of sharing/losing credit
  - researcher opposes early sharing before having a clear approach in mind or some results

- old printed paper model constrains the page #/format of publishable material
  - researcher opposes open collaboration due to fear of the others’ approach to the problem
  - researchers are pushed to publish as many full length papers as possible

- a lot of papers repackage the same ideas, resulting in dissemination overhead
  - dissemination of knowledge is limited/late
  - evolvement of knowledge is limited/slowed down

- authors waste unnecessary efforts on repackaging ideas
  - reviewing a paper becomes a challenge with all this overhead
  - consumers are frustrated when it comes to selecting papers/researchers

- unfair credit attribution (for authors, papers, reviewers, etc.)
What we plan to focus on

- researcher opposes early sharing due to fear of sharing/losing credit
- researcher opposes early sharing before having a clear approach in mind or some results
- researcher opposes open collaboration due to fear of the others’ approach to the problem
- researchers are pushed to publish as many full length papers as possible

- dissemination of knowledge is limited/late
- a lot of papers repackage the same ideas, resulting in dissemination overhead

- evolvement of knowledge is limited/slowed down
- authors waste unnecessary efforts on repackaging ideas
- reviewing a paper becomes a challenge with all this overhead

- consumers are frustrated when it comes to selecting papers/researchers
- unfair credit attribution (for authors, papers, reviewers, etc.)

reviewers may be biased or lack knowledge in a specific field

very competitive system for gaining reputation

old printed paper model constrains the page #/format of publishable material

What we plan to focus on

authors waste unnecessary efforts on repackaging ideas very competitive system for gaining reputation old printed paper model constrains the page #/format of publishable material
Architectural Model
Architectural Model

Attributes:
- title
- uri
- date
- body
- visibility
- copyright
...

Attributes:
- name
- affiliation
- address
- email
- url
...

researcher

SKO
Architectural Model

- Problem Definition
- LP System
  - Framework
  - Layer Cake
- LPrep Module
  - For Research
  - For Researchers
  - For Other Entities
  - Applications
- Conclusion

Architectural Model Diagram:
- **researcher**
  - coauthors
    - affiliated_with
    - same_board_as
    - depend
    - collab_compete
  - owns downloaded
  - reviewed
- **SKO**
  - part_of variation_of
  - review_of
  - cites
  - submitted_to

Legend:
- Light gray: explicitly declared objective relations
- Medium gray: implicitly declared objective relations
- Dark gray: subjective relations
LP Layer Cake

- Problem Definition
- LP System
  - Framework
  - Layer Cake
- LPrep Module
  - For Research
  - For Researchers
  - For Other Entities
  - Applications
- Conclusion
LP Layer Cake

Organisational Charters

LP System
- Integrity Constraints
- Nodes
- Relations

LP Reputation Service

LP Process Modelling Service

Problem Definition

LP System
- Framework
- Layer Cake

LPrep Module
- For Research
- For Researchers
- For Other Entities
- Applications

Conclusion
This layering allows the user to create complex processes for creating, sharing, and publishing knowledge objects; it also allows the user to calculate personalised reputation measures of varied complexities/applications.

The system imposes minimal (integrity) constraints on how to create, share, and publish knowledge objects; allowing these processes to easily evolve with time, as needed.
What issues does this framework solve?

- **Problem Definition**
  - **LP System**
    - Framework
    - Layer Cake
  - **LPrep Module**
    - For Research
    - For Researchers
    - For Other Entities
    - Applications
  - **Conclusion**

authors waste unnecessary efforts on repackaging ideas
very competitive system for gaining reputation
old printed paper model constrains the page #/format of publishable material
reviewers may be biased or lack knowledge in a specific field
researcher opposes early sharing due to fear of sharing/losing credit
researcher opposes early sharing before having a clear approach in mind or some results
researcher opposes open collaboration due to fear of the others’ approach to the problem
researchers are pushed to publish as many full length papers as possible
a lot of papers repackage the same ideas, resulting in dissemination overhead
authors waste unnecessary efforts on repackaging ideas
reviewing a paper becomes a challenge with all this overhead
consumers are frustrated when it comes to selecting papers/researchers
unfair credit attribution (for authors, papers, reviewers, etc.)
Reputation of a SKO is based on opinions & citations:

\[ SKO_{rep}(n) = Crep(n)^{\gamma_1} \cdot Orep(n)^{\gamma_2} \] (1)
We consider citations to propagate upwards in a graph.

Algorithm 1 Propagation of the citation of SKO node $n_2$

```plaintext
if $\text{record}(\text{cites}(n_1, n_2))$ then
    $\mathcal{N} = \{x \mid \text{compiles}(x, n_2)\}$
    $\mathcal{N}_i = \emptyset$
    $d = \varsigma$
    \textbf{while } $\mathcal{N} \neq \emptyset \land d > 0$ \textbf{do}
        \textbf{for } $n_i \in \mathcal{N}$ \textbf{do}
            \textbf{if } cites$(n_1, n_i) \notin C$ \textbf{then}
                $C = C \cup \{\text{cites}(n_1, n_i)\}$
                $\mathcal{N}_i = \mathcal{N}_i \cup \{n_i\}$
            \textbf{end if}
        \textbf{end for}
        $\mathcal{N} = \{x \mid \forall n_i \in \mathcal{N}_i \cdot \text{compiles}(x, n_i)\}$
        $\mathcal{N}_i = \emptyset$
        $d = d - 1$
    \textbf{end while}
\textbf{end if}
```
Citation Based Reputation

Instead of simply counting the number of citations a node has, a ranking algorithm may be used to describe the importance (or degree of authority) a node has by analysing which nodes links to which other.

The rank of node $n_i$ is the aggregation of the rank of the nodes linking to $n_i$. Similarly, the rank of $n_i$ will flow to those nodes $n_i$ links to, or cites.

PageRank is based on this idea:
Citation Based Reputation

Instead of simply counting the number of citations a node has, a ranking algorithm may be used to describe the importance (or degree of authority) a node has by analysing which nodes links to which other.

The rank of node $n_i$ is the aggregation of the rank of the nodes linking to $n_i$. Similarly, the rank of $n_i$ will flow to those nodes $n_i$ links to, or cites.

PageRank is based on this idea:

$$PR(n) = \frac{1 - d}{N} + d \sum_{\forall n_i \in M(n)} \frac{PR(n_i)}{L(n_i)}$$

where, $M(n)$ is the set of nodes linking to $n$, $L(n_i)$ is the number of outbound links on node $n_i$, $N$ is the total number of nodes, and $d$ is the damping factor.
Opinion Based Reputation

Opinion Definition: $opinion(\alpha, n, pd, c)$
where, $\alpha$ is the reviewer,
$n$ is the SKO node being reviewed,
$pd$ is a probability distribution describing $\alpha$’s opinion in $n$,
and $c$ describes $\alpha$’s confidence in its opinion.
Opinion Based Reputation

Opinion Definition: \( \text{opinion}(\alpha, n, pd, c) \)
where, \( \alpha \) is the reviewer,
\( n \) is the SKO node being reviewed,
\( pd \) is a probability distribution describing \( \alpha \)'s opinion in \( n \),
and \( c \) describes \( \alpha \)'s confidence in its opinion.

We believe opinions propagate up & down a graph.
Opinion Based Reputation

Opinion based reputation is then calculated from the probability distribution $\mathbb{P}$ that takes into account the propagation and aggregation of opinions:

$$\mathbb{P}(n) = \mathbb{P}_{\text{assigned}}(n)^{\xi_1} \cdot \mathbb{P}_{\text{inferred}}(n)^{\xi_2}$$  \hspace{1cm} (2)

$$\mathbb{P}_{\text{assigned}}(n) = \kappa \cdot \sum_{\forall \text{opinion}_a(\alpha,n,pd,c) \in O} r(\alpha, n) \cdot c \cdot pd \hspace{1cm} (3)$$

$$\mathbb{P}_{\text{inferred}}(n) = \kappa' \cdot \sum_{\forall \text{opinion}_i(\alpha,n,pd,c) \in O} r(\alpha, n) \cdot pd \hspace{1cm} (4)$$

How is $r(\alpha, n)$ calculated? How are inferred opinions ($\text{opinion}_i(\alpha,n,pd,c)$) calculated?
Opinion based reputation is then calculated from the probability distribution $\mathbb{P}$ that takes into account the propagation and aggregation of opinions:

$$\mathbb{P}(n) = \mathbb{P}_{\text{assigned}}(n)^{\xi_1} \cdot \mathbb{P}_{\text{inferred}}(n)^{\xi_2}$$ (2)

$$\mathbb{P}_{\text{assigned}}(n) = \kappa \cdot \sum_{\forall \text{opinion}_a(\alpha, n, pd, c) \in O} r(\alpha, n) \cdot c \cdot pd$$ (3)

$$\mathbb{P}_{\text{inferred}}(n) = \kappa' \cdot \sum_{\forall \text{opinion}_i(\alpha, n, pd, c) \in O} r(\alpha, n) \cdot pd$$ (4)

How is $r(\alpha, n)$ calculated?
How are inferred opinions ($\text{opinion}_i(\alpha, n, pd, c)$) calculated?
Opinion Based Reputation

Opinion based reputation is then calculated from the probability distribution $P$ that takes into account the propagation and aggregation of opinions:

$$P(n) = P_{\text{assigned}}(n)^{\xi_1} \cdot P_{\text{inferred}}(n)^{\xi_2}$$  \hspace{1cm} (2)

$$P_{\text{assigned}}(n) = \kappa \cdot \sum_{\forall \text{opinion}_a(\alpha, n, pd, c) \in O} r(\alpha, n) \cdot c \cdot pd$$  \hspace{1cm} (3)

$$P_{\text{inferred}}(n) = \kappa' \cdot \sum_{\forall \text{opinion}_i(\alpha, n, pd, c) \in O} r(\alpha, n) \cdot pd$$  \hspace{1cm} (4)

How is $r(\alpha, n)$ calculated? (see next)
How are inferred opinions ($\text{opinion}_i(\alpha, n, pd, c)$) calculated?
Propagation of Opinions

- Problem Definition
- LP System
  - Framework
  - Layer Cake
- LPrep Module
  - For Research
  - For Researchers
  - For Other Entities
  - Applications
- Conclusion

- a node with an assigned opinion
- a node with an inferred opinion
Propagation of Opinions
Propagation of Opinions

![Diagram showing the propagation of opinions with nodes and arrows]

- **Problem Definition**
- **LP System**
  - Framework
  - Layer Cake
- **LPrep Module**
  - For Research
  - For Researchers
  - For Other Entities
  - Applications
- **Conclusion**

- A node with an assigned opinion
- A node with an inferred opinion
- A node with a newly assigned opinion
Propagation of Opinions

The graph is analysed by viewing the red node to be at its centre.
Since opinions propagate both upwards & downwards, then the direction of the links no longer matter.
Propagation of Opinions

The rings illustrate the minimum distance between any given node and the central one.
Nodes with inferred opinions are then updated one by one, starting with those with the least distance from the central node.

The update stops either when no more neighbours have been updated, or when the difference between the old & new inferred opinions is smaller than a predefined threshold.
Propagation of Opinions
Propagation of Opinions
Propagation of Opinions
Propagation of Opinions
Propagation of Opinions
Propagation of Opinions

Problem Definition

LP System
- Framework
- Layer Cake

LPrep Module
- For Research
- For Researchers
- For Other Entities
- Applications

Conclusion
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- author
- reviewer
- future collaborator
- invited speaker
- ...

Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- author
  - aggregation of the researcher’s SKOs' reputation
  - h-index based measures *
- reviewer
- future collaborator
- invited speaker
- ...

* h-index based measures indicate the number of papers that have been cited at least h times.
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- **author**
  - aggregation of the researcher’s SKOs’ reputation
  - h-index based measures *

- **reviewer**
  - history of bias
  - history of being right
  - history of being stubborn
  - reviewer’s social relations with others helps highlight possible dependency in opinion, and collaborative/competitive behaviour
  - effect of the social network on the researcher’s opinion

- **future collaborator**

- **invited speaker**

- ...
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- author
  - aggregation of the researcher’s SKOs’ reputation
  - h-index based measures *

- reviewer
  - history of bias
  - history of being right
  - history of being stubborn
  - reviewer’s social relations with others helps highlight possible dependency in opinion, and collaborative/competitive behaviour
  - effect of the social network on the researcher’s opinion
    - the opinions of the reviewer’s neighbours
    - the interconnectivity parameter of the social network
    - pinpointing the group leader

- future collaborator

- invited speaker

- ...

* h-index based measures refer to a metric in bibliometrics that attempts to measure the productivity and impact of a researcher by considering the number of their publications and the number of citations each of those publications receives.
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- **author**
  - aggregation of the researcher’s SKOs’ reputation
  - h-index based measures *

- **reviewer**
  - history of bias
  - history of being right
  - history of being stubborn
  - reviewer’s social relations with others helps highlight possible dependency in opinion, and collaborative/competitive behaviour
  - effect of the social network on the researcher’s opinion
    - the opinions of the reviewer’s neighbours
    - the interconnectivity parameter of the social network
    - pinpointing the group leader

- **future collaborator**
  - researcher’s centrality measure in his social network (degree, betweenness, closeness, prestige, etc.)

- **invited speaker**

- ...
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- **author**
  - aggregation of the researcher’s SKOs’ reputation
  - h-index based measures *

- **reviewer**
  - history of bias
  - history of being right
  - history of being stubborn
  - reviewer’s social relations with others helps highlight possible dependency in opinion, and collaborative/competitive behaviour
  - effect of the social network on the researcher’s opinion
    - the opinions of the reviewer’s neighbours
    - the interconnectivity parameter of the social network
    - pinpointing the group leader

- **future collaborator**
  - researcher’s centrality measure in his social network (degree, betweenness, closeness, prestige, etc.)

- **invited speaker**
  - the measure of interest of a person’s research

- ...
Reputation of Authors

A researcher’s reputation is dependent on the role. It may be affected by:

- **author**
  - aggregation of the researcher’s SKOs’ reputation
  - h-index based measures *

- **reviewer**
  - history of bias
  - history of being right
  - history of being stubborn
  - reviewer’s social relations with others helps highlight possible dependency in opinion, and collaborative/competitive behaviour
  - effect of the social network on the researcher’s opinion
    - the opinions of the reviewer’s neighbours
    - the interconnectivity parameter of the social network
    - pinpointing the group leader

- **future collaborator**
  - researcher’s centrality measure in his social network (degree, betweenness, closeness, prestige, etc.)

- **invited speaker**
  - the measure of interest of a person’s research

- ...

Of course, many roles share common points!
Reputation of Authors

Numerous methods may be followed when considering an author’s citation based reputation:

Citation count: \( \sum_{i=1}^{n} c_i \), where \( c \) is the citation’s count

Average citations per paper: \( \frac{1}{n} \sum_{i=1}^{n} c_i \), where \( c \) is the citation’s count

\( h \)-index: Given a set of papers ranked in decreasing order of the number of citations that they received, the \( h \)-index is the number \( h \) such that each of the top \( h \) papers have received at least \( h \) citations while the remaining papers have received at most \( h \) citations.

\( g \)-index: Given a set of papers ranked in decreasing order of the number of citations that they received, the \( g \)-index is the (unique) largest number such that the top \( g \) papers received (together) at least \( g^2 \) citations.

\( m \)-index: \( \frac{h}{y} \), where \( h \) is the \( h \)-index, and \( y \) is the number of years since publishing the first paper

\( a \)-index: \( \frac{1}{h} \sum_{j=1}^{h} c_j \), where \( h \) is the \( h \)-index, and \( c \) the citations count
Reputation of Other Knowledge Entities

- a conference
- a publisher
- a library
- ...

These are usually computed based on the reputation of SKOs they include, publish, archive, etc.
Useful Applications

- finding an appropriate future collaborator
- fishing for new SKOs for a given conference
- deciding how many SKOs per field to accept (based on a minimum accepted quality)
- finding reputable PC members / reviewers
- deciding who to review an SKO
- deciding how many reviewers per SKO are needed
- aggregating reviewers results
- ...

A Conference Process

declare a conference series
place a call for papers
review submitted papers
select accepted papers
invite selected papers
declare a conference instance
publish proceedings
...
What issues does the LPrep module solve?

- Researchers oppose early sharing due to fear of sharing/losing credit.
- Researchers oppose early sharing before having a clear approach in mind or some results.
- Researchers oppose open collaboration due to fear of the others' approach to the problem.
- Researchers are pushed to publish as many full length papers as possible.
- A lot of papers repackaged the same ideas, resulting in dissemination overhead.
- Dissemination of knowledge is limited/late.
- Evolvement of knowledge is limited/slowed down.
- Authors waste unnecessary efforts on repackaging ideas.
- Reviewing a paper becomes a challenge with all this overhead.
- Consumers are frustrated when it comes to selecting papers/researchers.
- Unfair credit attribution (for authors, papers, reviewers, etc.).

- Very competitive system for gaining reputation.
- Old printed paper model constrains the page #/format of publishable material.
- Reviewers may be biased or lack knowledge in a specific field.
- Authors waste unnecessary efforts on repackaging ideas.
- Consumers are frustrated when it comes to selecting papers/researchers.
For Further Information:

http://project.liquidpub.org/