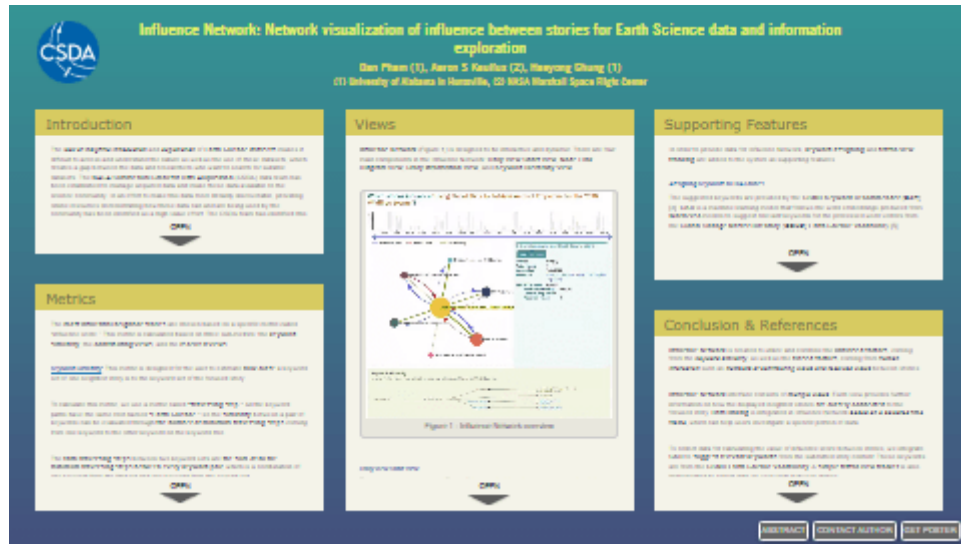


Influence Network: Network visualization of influence between stories for Earth Science data and information exploration



Dan Pham (1), Aaron S Kaulfus (2), Haeyong Chung (1)

(1) University of Alabama in Huntsville, (2) NASA Marshall Space Flight Center

PRESENTED AT:



INTRODUCTION

The *lack of insightful introduction and explanation* of **Earth Science datasets** makes it difficult to access and understand the nature as well as the use of these datasets, which creates a gap between the data and researchers who want to search for suitable datasets. We, the **NASA Commercial Smallsat Data Acquisition** (CSDA) data team, has been established to evaluate and manage **highly-potential datasets** [1] as well as make these data **available** and **broadly discoverable** to the science community.

To address this issue, we present the **“Data in Action” (DIA) Story** system [2, 3] to provide tools for sharing data in the form of stories. Particularly, the **DIA Story** system employs **data storytelling** which is an approach combining multimedia, data, and storytelling to deliver insightful stories centered around data.

Influence Network is a frontend component that is a part of the data exploration tools within the **DIA Story** system. It assist users and story authors in *discovering similar stories* through **influence flows** as well as *studying reader behavior and the currently popular phenomenon* in the field of **Earth Science**.

In **Influence Network**, there are *two types* of stories: the **focused story**, which is the story that a user is currently reading, and **neighbor stories**, which are other stories within the DIA Story system that have the most influence on the focused story.

Influence Network provides:

- **Interacting** with **dynamic node-link diagram** supported with **force-directed graph layout**
- **Filtering** displayed data based on the **selected time frame** from the Time Slider
- **Switching** Daily View Chart mode between **"stacked"** and **"grouped"**
- **Searching** for stories **based on keywords presented** in the Keyword Hierarchy View
- Allowing users to **redirect** to the individual story page of a displayed neighbor story from the Story Information View

METRICS

The **most influential neighbor stories**, which are stories recommended by Influence Network, are chosen based on a specific metric called “influence score.” This metric is calculated based on three submetrics: the **keyword similarity**, the **contributing views**, and the **received views**. These submetrics combine **user input** with **objective measures of similarity**. This will be further discussed in the section “**Supporting Features**.”

Keyword Similarity: This metric is designed for the user to estimate **how close** a keyword set of one neighbor story is to the keyword set of the focused story.

To calculate this metric, we use a metric called “**traversing step**.” All the keyword paths have the same root named “**Earth Science**,” so the **similarity** between a pair of keywords can be evaluated through **the number of minimum traversing steps** coming from one keyword to the other keyword on the keyword tree.

The **total traversing steps** between two keyword sets are **the sum of all the minimum traversing steps between every keyword pair**, which is a combination of one keyword from the first set and one keyword from the second set.

The **smaller** the total traversing steps are, the **closer** the two keyword sets. Therefore, after calculating the list of the traversing steps between the focused story and other stories in the DIA Story system, we need to **reverse the order of all traversing steps** between the two keyword sets of the focused story and other stories in the DIA Story system. By doing so, the smallest value is transformed into the highest value, and vice versa.

The final value is **normalized** by being transformed into **percentage**.

Contributing View Score: This metric estimates the influence of one neighbor story to the focused story through **the number of traffic views that come from this neighbor story to the focused story**.

The value of this metric is straightforward since it is related to the number of traffic views **without having to be pre-processed** like the *keyword similarity*. Similar to the *keyword similarity*, the **contributing view score** is **transformed** into **percentage** before it is combined into the final influence score.

Received View Score: This metric estimates the influence of one neighbor story to the focused story through **the number of traffic views that come from the focused story to this neighbor story**.

The **received view score** is calculated in the same way as the **contributing view score**.

Influence Score: The influence score estimates the **overall influence** of one neighbor story to the focused story. It is used to determine **the most influential neighbor stories** of a focused story.

It is the **combination** of the **keyword similarity**, the **contributing view score**, and the **received view score**, and is calculated by **taking the average** of these scores.

VIEWS

Influence Network (Figure 1) is designed to be interactive and dynamic. There are four main components in the Influence Network: **Daily View Chart View**, **Node-Link Diagram View**, **Story Information View**, and **Keyword Hierarchy View**.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1669931340/agu%20fm2022/5b-71-59-61-70-b1-b2-8b-ee-d9-00-5a-d4-ba-b0-c0/image/influence_network_overview_cb2zvm.mp4

Figure 1 - Influence Network overview

Daily View Chart View

The **Daily View Chart View** (Figure 2) is a vertical bar chart that presents the daily number of traffic views contributing to the focused story coming from its most influential neighbor stories and the other sources. These sources can be neighbor stories that are not displayed in Influence Network, home page, search result page, or external sources.

These contributing views can show the **reading frequency** of DIA Story users, which can assist story authors and researchers in **learning reader behavior**. Furthermore, a part of the contributing views are from the displayed neighbor stories in the Node-Link Diagram and are also a **contributing factor** of the **influence score** of a neighbor story to the focused story.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1669930151/agu%20fm2022/5b-71-59-61-70-b1-b2-8b-ee-d9-00-5a-d4-ba-b0-c0/image/stacked_grouped_nfrbam.mp4

Figure 2 - The interface of the Daily View Chart View

The bars in the **Daily View Chart View** are displayed in chronological order.

It is possible to switch the presentation of the *Daily View chart* between “**grouped**” and “**stacked**” by clicking on the corresponding radio button, which is positioned above the chart.

By hovering over a specific neighbor story node in the **Node-Link Diagram View**, the views coming from the respective story are highlighted while the rest is grayed out.

Time Slider is one of the main inputs in the **Influence Network**. It *lines up with the x-axis* of the Daily View Chart which represents the time (date) that the views were recorded.

It is a *horizontal slider* which has *two handles*, presenting the start date and end date of the selected time frame.

Users can interact with the **Time Slider** by sliding the two handles along the horizontal line to *extract a portion of the displayed data*.

Node-Link Diagram View

In this view, each **node represents a story** within the DIA Story system, and each **link represents a relationship** between a displayed neighbor story and the focused story (Figure 3).

This view promotes the **quantitative flows** of the **three contributing metrics** of the **influence score: contributing view, receiving view, and keyword similarity**. Therefore, it can provide more insight on how the displayed neighbor stories **are closely connected** to the focused story.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1669930787/agu%20fm2022/5b-71-59-61-70-b1-b2-8b-ee-d9-00-5a-d4-ba-b0-c0/image/influence_network_zxaizu.mp4

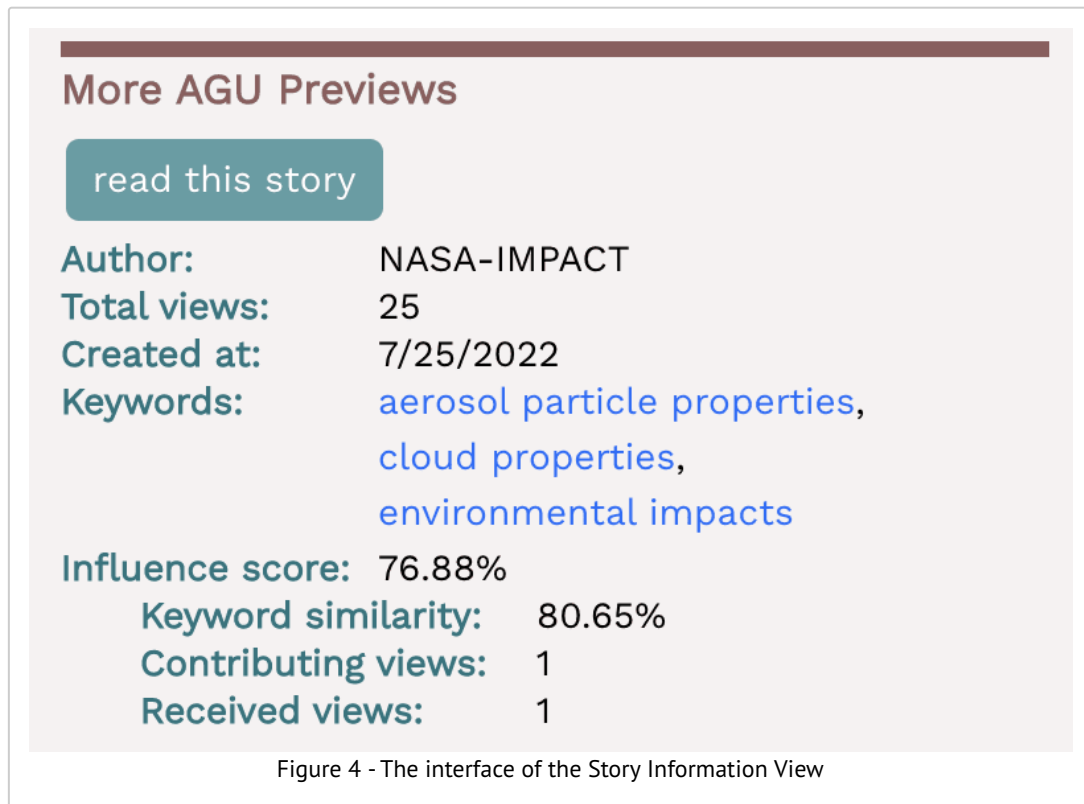
Figure 3 - The interface of the Node-Link Diagram View

The **higher** the **keyword similarity score** between the two stories, the **thicker** the **link** that connects them. The thickness of the links can also change **based on a selected time frame** that users want to investigate.

The implementation of this diagram is supported with the **force-directed graph layout**.

Story Information View

Story Information View (Figure 4) contains the basic information of the most recent hovered story from the **Node-Link Diagram View**. Additionally, when hovering on a neighbor story, the details of the **influence score** between these stories is show, *providing more information on how they are connected* to each other.



The assigned keywords consist of **clickable keywords**. When users click on one of these keywords, they will be switched to the search result page that filters and displays all stories related to the selected keyword.

A “**read this story**” button is also implemented in the **Story Information View** to provide the URL that redirects users to the individual page of the hovered story.

The information in this view can be changed dynamically by switching the hovered target from the **Node-Link Diagram View**.

Keyword Hierarchy View

Since all of the presented keywords have a common root node, the keyword sets of the focused story and the most recent hovered neighbor story can be presented as **one single tree**. The **Keyword Hierarchy View** (Figure 5) visualizes this tree, which can provide further details on how the **keyword similarity score** is calculated.

[VIDEO] https://res.cloudinary.com/amuze-interactive/image/upload/f_auto,q_auto/v1669931027/agu%20fm2022/5b-71-59-61-70-b1-b2-8b-ee-d9-00-5a-d4-ba-b0-c0/image/keyword_hierarchy_sawxih.mp4
Figure 5 - The interface of the Keyword Hierarchy View

Similar to the **Story Information View**, these keywords are **clickable**, which helps users to filter stories with a specific keyword.

The information in this view can also be changed dynamically by switching the hovered target from the **Node-Link Diagram View**.

SUPPORTING FEATURES

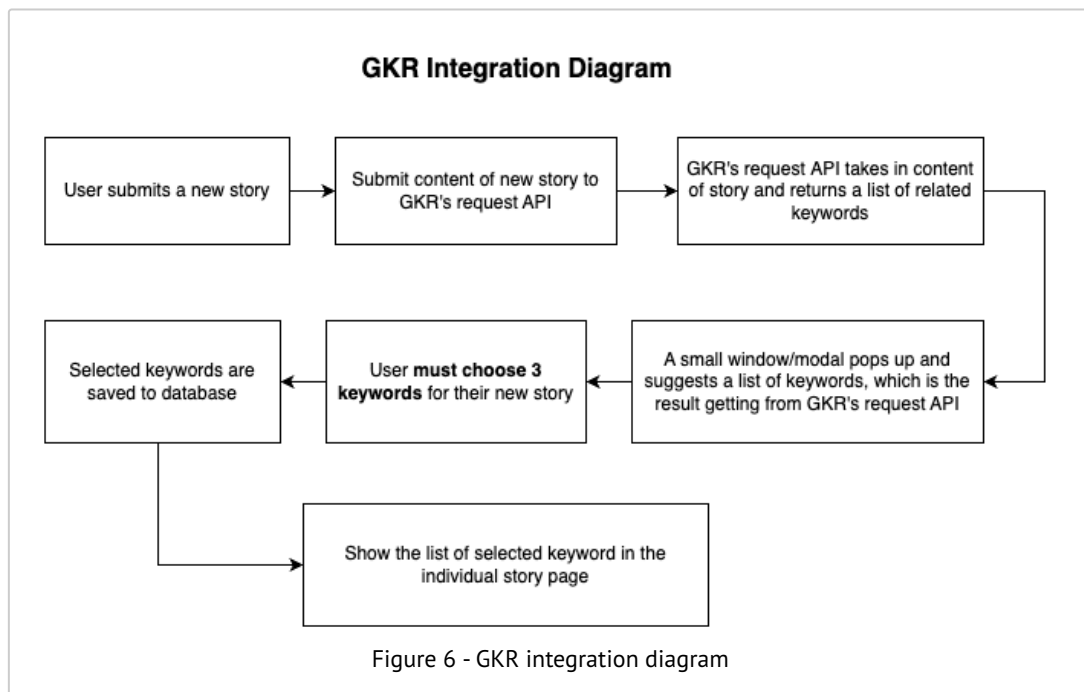
In order to provide data for Influence Network, **keyword assigning** and **traffic view tracking** are added to the system as supporting features.

Assigning Keywords to DIA Stories

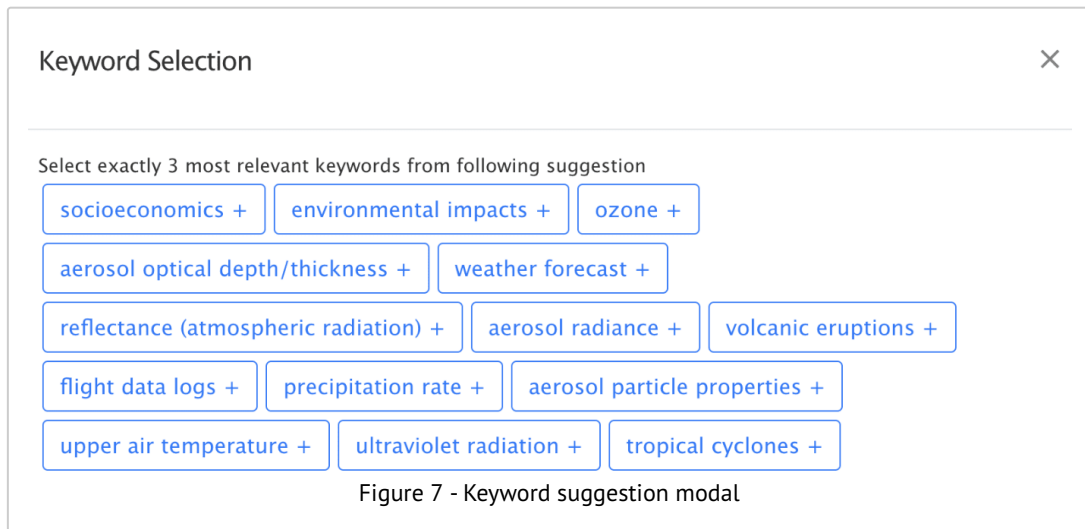
The suggested keywords are provided by the **GCMD Keyword Recommender (GKR)** [4]. **GKR** is a machine learning model that follows the word embeddings produced from **Word2Vec** models to suggest relevant keywords for the processed word vectors from the **Global Change Master Directory (GCMD) Earth Science vocabulary** [5].

After sending the story content to the GKR using its API, the output would contain the keywords that are related to the submitted content.

The process of GKR integration is described in Figure 6.



When the story author submits their story for review, they will be asked to select relevant keywords for their story. The suggested keywords are displayed in the popped-up modal (Figure 7). The author needs to select exactly three keywords before they finish submitting the story for review.



Tracking Total Views and Traffic Views Between DIA Stories

To track each view on a specific story, a JavaScript function will be called every time the Document Object Model (DOM) is ready, and an ajax call is created to call an API from the Django backend in order to record a new view on the currently reading story.

Currently, new traffic views between stories in the DIA Story system are tracked based on specific hyperlinks and buttons. An ajax call to the traffic view record API is made before users are redirected to the next story to record this new traffic view.

CONCLUSION & REFERENCES

Influence Network is created to utilize and combine the **unbiased factors**, coming from the *keyword similarity*, as well as the **biased factors**, coming from *human interaction* such as *numbers of contributing views and received views* between stories.

Influence Network interface consists of *multiple views*. Each view provides further information on how the displayed neighbor stories **are closely connected** to the focused story. **Data linking** is integrated in Influence Network *based on a selected time frame*, which can help users investigate a specific portion of data.

To collect data for calculating the value of influence score between stories, we integrate **GKR** to **suggest relevant keywords** from the submitted story content. These keywords are from the **GCMD Earth Science vocabulary**. A **simple traffic view tracker** is also implemented to collect data on user visits between stories.

We believe **Influence Network** can bring up possibilities for users, specifically story authors, to learn about **reader behavior** and the **currently popular phenomenon** in the field of **Earth Science**.

References

[1] Manil Maskey et al. **Commercial Smallsat Data Acquisition (CSDA) Program Update**. July 2021. url: <https://ntrs.nasa.gov/citations/20210018254>
(<https://ntrs.nasa.gov/citations/20210018254>)

[2] Aaron Kaulfus et al. **Data System for NASA's Commercial Smallsat Data Acquisition Program**. Dec. 2021. url: <https://ntrs.nasa.gov/citations/20210025632>
(<https://ntrs.nasa.gov/citations/20210025632>)

[3] Aaron Kaulfus et al. **NASA's Commercial Smallsat Data Acquisition Program Data Stewardship and Data Management**. Jan. 2022. url: <https://ntrs.nasa.gov/citations/20220000472>
(<https://ntrs.nasa.gov/citations/20220000472>)

[4] NASA's Interagency Implementation and Advanced Concepts Team. **GCMD Keyword Classifier**.

[5] Earth Science Projects Division, Goddard Space Flight Center (GSFC). **Global Change Master Directory (GCMD)**. url: <https://forum.earthdata.nasa.gov/app.php/tag/GCMD+Keywords>
(<https://forum.earthdata.nasa.gov/app.php/tag/GCMD+Keywords>)

ABSTRACT

Using storytelling to present data has been demonstrated as an effective way to help data users gain deeper insight of information. For this purpose, the “Data in Action” story concept has been adopted by NASA Commercial Smallsat Data Acquisition (CSDA) Program to encourage creating and sharing data information in story form. To aid researchers in exploring similar stories and data in their fields of interest, we created an “Influence Network” within the “Data in Action” framework. The Influence Network is a data visualization system component which gathers information of story relationships using a concept called “influence,” which we define based on the number of visits and keyword similarity between stories. The visualization then provides insights into the influence flows between different stories within the system. The visualization of this “Influence Network” focuses on one story at a time, introducing a time series of neighbor stories that have the most influence to the currently focused story. By focusing on views over time and visualizing influence flows between stories, we aim to assist authors in understanding how their readers perceive their stories as well as advancing their methods for delivering more meaningful stories to expand and lower the barrier to use of CSDA datasets.

