WORKFLOW DESCRIPTION

MAPPING THE CO-EVOLUTION OF ARTIFICIAL INTELLIGENCE, ROBOTICS, AND INTERNET OF THINGS OVER 20 YEARS (1998-2017)

Compiled by Tenzin Choeden, Olga Scrivner

CRANE CONVERGENCE NETWORK WORKFLOW

1. PURPOSE

The purpose of this workflow document is to make our research reproducible. Here we outline the steps we used to create the convergence network (Fig.10) of our research paper (Mapping the Co-Evolution of Artificial Intelligence, Robotics, and Internet of Things Over 20 Years (1998-2017).

2. TOOLS

Here are the list of tools stacks we used for creating the convergence network.

i. Jupyter Notebook (python)
ii. Gephi
iii. Adobe Illustrator

3. DATA SOURCES

The data we used here comes from two main sources.

i. **Web of Science database:**
We extracted data consisting of two columns, WOS_id and WOS_ref from Web of Science database. Each row represents the source and target of references. WOS_id and WOS_ref consist of unique identification number assigned to a paper. We extracted three csv files, corresponding to each field of interest (AI,

IoT and Robotics). See WoS relational database schema - Link: https://cadre.iu.edu/resources/documentation/wos_core_erd%20(1).pdf

ii. **Web of Science Online Portal:**

We used Institute of Scientific Information (ISI) publication data from Web of Science Online Portal to get the year and other relevant descriptions of papers. ISI files are converted to CSV using Make-a-Vis available at https://make-a-vis.netlify.com/

4. PROCESS

The process is broken down into three important steps:

i. **Data Preprocessing**

The raw data from the data sources mentioned above was preprocessed and transformed into a required format. The preprocess procedure is described in detail with code in the python notebook available on https://github.com/cns-iu/AICoEvolution

The required node output needs to be in the form of:

Id	Label	Lat	Lon
1	2014	40	160
2	2014	40	120

The latitude and longitude values are used to arrange the nodes in 2D space in Gephi.

The edge output looks like this:

source	target	weight	cat
1	2	1	0
3	4	1	1

The 'cat' value corresponds to the source publication's topic. It is used in Gephi to determine the color of the line (0 = IoT, 1 = Robotics, 2 = AI). The 'weight'

value is used to determine the number of citations from one year to another, between two topics.

ii. Network graph using Gephi

Once we get the required file format for both edge and node of graph, we used Gephi to create the network visualization. We have to make sure we import the geo-layout plugin before proceeding with the steps below.

a. Import the node and edge csv files as spreadsheet. Note: Select append to existing workspace when you import the second file
b. Treat the file as a geospatial visualization, run the 'Geo Layout' for the network layout
c. Set the edge ranking value to the 'cat' value from the edges file
d. Select colors for the low, middle, and high values. Corresponding to 0, 1, and 2, or the three topic areas. We used #29279F, #C51162, #FFD600 for the three colors.
e. In preview, deselect the curved option in Edges, Set the color – original
f. Adjust the network as you see fit and export to SVG format

iii. Final touch-up using Adobe Illustrator

Adobe Illustrator was used to add horizontal lines, adjust node sizes, add labels, add legends and adjust scaling to fit into the paper.

GEOSPATIAL VISUALIZATION WORKFLOW

1. PURPOSE

The purpose of this workflow document is to make our research reproducible. Here we outline the steps we used to create the geospatial visualization (Fig.8), of our research paper (Mapping the Co-Evolution of Artificial Intelligence, Robotics, and Internet of Things Over 20 Years (1998-2017).

2. TOOLS

Here are the list of tools stacks we used for creating the geospatial network.

i. Make-A-Viz - https://make-a-vis.netlify.com/

ii. Gephi
iii. Adobe Illustrator

3. DATA SOURCES

The data we used here comes from following source.

i. **Web of Science Online Portal:**
We used Institute of Scientific Information (ISI) publication data from Web of Science Online Portal with paper ids, title, authors, address, year of publication, number of citations.

4. PROCESS

The process is broken down into three important steps:

i. **Pre-processing with Make-a-Vis**
The input file is fed to Make-a-Vis here. From this we generated the required data fields and format that will be required for the next step. There are two output files from this process, co-author link and author coordinates. Convert these to files into nodes and edges for network visualization: add unique ID to each author in author coordinate file and create a source/target ids in co-author link file - see gephi-parser.ipynb.

ii. **Geospatial viz. using Gephi**
Once we get the required file format for both edge and node of graph, we used Gephi to create the network visualization. We have to make sure we import the geo-layout plugin before proceeding with the steps below.

1. Import files edges-ai.csv as Gephi edges and nodes-ai.csv as Gephi nodes
   - Make sure edges are undirected
2. Run Geo Layout
   - Select "Geo Layout" under "Layout" section
   - Make sure the projection is selected as "Mercator"
   - Click run
3. Filter non-null latitudes
   - Under the "Filters" section, expand "Attributes", "Non-null", then select "latitude"
   - Click filter
4. Copy filtered graph to new workspace
   - Under the filters section, click export button:
   - **Alternative:** Select the graph using the mouse/rectangle selection

- Exclude remote nodes (islands) outside of mainland U.S.
- Right click “Move to...” > “New Workspace”
- Skip step 5

5. Delete remote nodes (islands) outside of mainland U.S.
   - Using mouse tool, select remote node(s)
   - Right click > Delete
6. Adjust node color by date of first publication
   - Under “Appearances”, while the “Nodes” tab is highlighted, click size button:
   - Click “Ranking” tab
   - Select “First Year”
   - Hover over the color bar and double click the marker to set its color
     - Left marker hex:00695C
     - Middle marker hex: 26A69A
     - Right marker hex: E0F2F1
7. Adjust node size to citations (4-40 min/max)
   - Under “Appearances”, while the “Nodes” tab is highlighted, click size button:
   - Click “Ranking” tab
   - Select “Citations”
   - Set min:4, max:40
8. Filter edges by Joint Publication
   - Under the “Filters” section, expand “Attributes”, “Range”, then select “jointpublications”
   - Change the range to:
     - For AI: 2-13
     - For Robotic: 5-23
     - For IOT: 3-20
9. Export as SVG

iii. Final touch-up using Adobe Illustrator

Adobe Illustrator was used to overlay the resulting SVG, adjust node sizes, add labels, add legends and adjust scaling to fit into the paper.

1. Import a U.S. Mercator projection map
2. Import the SVG file
3. Map the cities from SVG file manually to fit the map

CO-AUTHOR NETWORK WORKFLOW

1. DATA SOURCES

The data we used here comes from following source.

i. **Web of Science Online Portal:**
We used Institute of Scientific Information (ISI) publication data from Web of Science Online Portal to get the year and other relevant descriptions of papers.

2. TOOLS

Here are the list of tools stacks we used for creating the geospatial network.

i. Sci2 - https://sci2.cns.iu.edu/user/index.php
ii. Gephi
iii. Adobe Illustrator

3. PROCESS

The process is broken down into two important steps:

i. **Preparing and creating Graphxml**
Create a Property text file for co-occurrence network with the content below:
node.numberOfWorks = Authors.count
node.firstPubYear = Year.min
node.timesCited = TC.sum
edge.numberOfCoAuthoredWorks = Authors.count
edge.firstPub = Year.min

1. Use iot-1998-2017.isi documents
2. Load to Sci2
3. Export as csv – extracted unique ISI records
    - Rename columns to fit isiCoAuthorship.properties file: Authors, Year, TC
    - Create a csv with just these three columns
4. Reload to Sci2
5. Select Data preparation>Extract co-occurrence network
    - Select properties>isiCoAuthorship.properties

6. Run network analysis – top and write stats
7. Save csv output
8. Remove zero timecited (TC >=1 for ai, TC >= 20 for Robotics, TC >=1 for IOT
    - Save as a filtered file
9. Load back to sci2
10. Remove isolates: Preprocessing – remove isolates (make sure to select extracted network file)
11. Run Analysis > network – weakly component and select the largest one
12. Save as graphml (peruse)
13. Load to Gephi Extracted network (graphml)
14. Run ForceAtlas2, select Prevent Overlap
15. Graph legend
    - Node size: # Citations
    - Node color: FirstPublishYear (scale)
    - Label size: # Co-Authors (degree) --> #Citation
    - Edge width: # Joint Publications

ii. Creating GEPHI network

To create the network visualization we used Gephi and the geo-layout plugin.

1. ForceLayout2 – Run (no overlap)
2. Change gravity from 1 to higher (e.g. 3) – Run
3. Appearance – Node > Size > Ranking > TimesCited >min and max [adjust as needed] > Apply
    - AI: min 4/ max 40
    - Rob: min 1/max 50
    - IoT: min 1/max 4
4. Appearance – Node > Color > Ranking > FirstPubYear > swipe color staring from darker to lighter > Apply
    - Color: - right dark – 00695C, middle 26A69A, left light - E0F2F1
5. Appearance – Node > T label size > Ranking > TimesCited > Apply > Select T on the bottom (all labels are shown)
6. Go to Filter panel –right: Filter > Attribute > Range > TimeCited [double click]. Filter as needed, select A-> icon next to the filter, then stop
7. Layout > label adjust Run
8. Preview: Edge straight, unselect proportional size, font size – 48, show labels, save as SVG and PNG