

# Detecting Influencers in Written Online Conversations

Or Biran

Sara Rosenthal

Jacob Andreas

Kathleen McKeown

Owen Rambow

# What is an Influencer?

- Someone whose opinions/ideas profoundly affect the conversation
- An influencer may have the following characteristics (Katz and Lazarsfeld, 1955)
  - alter the opinions of their audience
  - resolve disagreements where no one else can
  - be recognized by others as one who makes important contributions
  - often continue to influence a group even when not present
  - have other conversational participants adopt their ideas and even the words they use to express their ideas
- More formally, an influencer:
  - Has credibility in the group
  - Persists in attempting to convince others, even if some disagreement occurs
  - Introduces topics/ideas that others pick up on or support

# What an Influencer is not

- There are types of power which a participant in online conversations can have that are not indicative of him/her being an influencer
- **Hierarchical power**
  - A participant appears to be above others in a hierarchy. For example:
    - The participant appears to give an approval or a direct order
    - Another person appears to be asking the participant for approval
    - The participant appears to have authority to make the final decision
- **Situational power**
  - Someone appears to have power (authority to direct / approve other people's actions) in the current situation or while a particular task is being performed
- **Power directing the communication**
  - Someone actively attempts to achieve the goal of the discussion and directs communication towards that goal

# Motivation

- Detecting Influencers in dialog is useful for product advertisement, politics, and defense
  - An influential person can be targeted to advertise new products
  - An influential person can be targeted to promote a political candidate
  - An influential person in a suspicious conversation can be monitored for suspicious behavior

# Related Work

- Rienks (2007)
  - Detecting influencers in a corpus of conversations
  - Focuses on non-linguistic behavior
- Quercia et al. (2011)
  - Language use of influencers vs. other groups in Twitter

# Data

- Threads taken from:
  - Wikipedia Discussions (Talk pages)
  - LiveJournal blogs

# Annotated Threads	
Wikipedia	88
LiveJournal	245

- Threads have tree structure (direct responses)
- Annotation
  - Two undergraduate students of liberal arts
  - IAA (Cohen's Kappa) is .72
  - Number of influencers per thread was 0-2
    - 2 quite rare

# Conversational Behaviors

- Hypothesis: An influencer, or an influencer's conversational partner, is likely to engage in the following conversational behaviors:
  - Persuasion
  - Agreement/disagreement
  - Specific Dialog Patterns

# Language Uses

- For each of the three conversational behaviors, define several Language Uses

Conversational Behavior	Language Use
Persuasion	Attempt to Persuade
	Claims
	Argumentation
Agreement/Disagreement	Agreement
	Disagreement
Dialog Patterns	Initiative
	Investment
	Irrelevance
	Interjection
	Incitation



# Example Thread

- Three participants: Arcadian, JFW and Emmanuelm, edit the article on Lymphoma
  - Arcadian makes a suggestion; there is some agreement
  - Emmanuelm joins and challenges Arcadian's idea
  - A short discussion follows, and Arcadian defers to Emmanuelm's opinion
- Our annotators marked Emmanuelm as the influencer

# Example Thread

<p><b>P1 by Arcadian</b> <i>&lt;pc<sub>1</sub>&gt;</i>There seems to be a much better list at the National Cancer Institute than the one we've got.<i>&lt;/pc<sub>1</sub>&gt;</i><i>&lt;pa<sub>1</sub>&gt;</i>It ties much better to the actual publication (the same 11 sections, in the same order).<i>&lt;/pa<sub>1</sub>&gt;</i>I'd like to replace that section in this article. Any objections?</p>
<p><b>P2 by JFW</b> <i>&lt;pc<sub>2</sub>&gt;</i><i>&lt;a<sub>1</sub>&gt;</i>Not a problem.<i>&lt;/a<sub>1</sub>&gt;</i><i>&lt;/pc<sub>2</sub>&gt;</i>Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal</p>
<p><b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.</p>
<p><b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <i>&lt;pc<sub>3</sub>&gt;</i>I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histiocytic and mastocytic neoplasms.<i>&lt;/pc<sub>3</sub>&gt;</i><i>&lt;pa<sub>3</sub>&gt;</i>Remember, Wikipedia is meant to be readable <i>&lt;/pa<sub>3</sub>&gt;</i>by all. Let me know what you think before I delete the non-lymphoma parts.</p>
<p><b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?</p>
<p><b>P6 by Emmanuelm</b> <i>&lt;d<sub>1</sub>&gt;</i>Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.<i>&lt;/d<sub>1</sub>&gt;</i>Would you mind moving the leukemia portion to the leukemia page?</p>
<p><b>P7 by Emmanuelm</b> <i>&lt;pc<sub>4</sub>&gt;</i>Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.<i>&lt;/pc<sub>4</sub>&gt;</i>My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.</p>
<p><b>P8 by Arcadian</b> <i>&lt;pa<sub>5</sub>&gt;</i><i>&lt;a<sub>2</sub>&gt;</i>To simplify the discussion, I have restored that section to your version.<i>&lt;/a<sub>2</sub>&gt;</i><i>&lt;/pa<sub>5</sub>&gt;</i>You may make any further edits, and <i>&lt;pc<sub>6</sub>&gt;</i>I will have no objection.<i>&lt;/pc<sub>6</sub>&gt;</i></p>
<p><b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <i>&lt;pc<sub>7</sub>&gt;</i>It would be defensible to list ALL and CLL here.<i>&lt;/pc<sub>7</sub>&gt;</i><i>&lt;pa<sub>7</sub>&gt;</i>They fall under the lymphoproliferative disorders.<i>&lt;/pa<sub>7</sub>&gt;</i></p>

# Persuasion

- An **attempt to persuade** is an expression of opinion (a claim) followed by explicit justification of the claim (an argumentation)
  - *Persuade to believe, not persuade to act*

**CLAIM:** There seems to be a much better list at the National Cancer Institute than the one we've got.

**ARGUMENTATION:** It ties much better to the actual publication (the same 11 sections, in the same order).

Example of Persuasion

- In addition, **claims** and **argumentation** are detected independently of one another

# Persuasion in the Example Thread

<p><b>P1 by Arcadian</b> &lt;pc<sub>1</sub>&gt;There seems to be a much better list at the National Cancer Institute than the one we've got.&lt;/pc<sub>1</sub>&gt;&lt;pa<sub>1</sub>&gt;It ties much better to the actual publication (the same 11 sections, in the same order).&lt;/pa<sub>1</sub>&gt; I'd like to replace that section in this article. Any objections?</p>
<p><b>P2 by JFW</b> &lt;pc<sub>2</sub>&gt;&lt;a<sub>1</sub>&gt;Not a problem.&lt;/a<sub>1</sub>&gt;&lt;/pc<sub>2</sub>&gt;Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal</p>
<p><b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.</p>
<p><b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! &lt;pc<sub>3</sub>&gt;I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histiocytic and mastocytic neoplasms.&lt;/pc<sub>3</sub>&gt;&lt;pa<sub>3</sub>&gt;Remember, Wikipedia is meant to be readable &lt;/pa<sub>3</sub>&gt;by all. Let me know what you think before I delete the non-lymphoma parts.</p>
<p><b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?</p>
<p><b>P6 by Emmanuelm</b> &lt;d<sub>1</sub>&gt;Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.&lt;/d<sub>1</sub>&gt;Would you mind moving the leukemia portion to the leukemia page?</p>
<p><b>P7 by Emmanuelm</b> &lt;pc<sub>4</sub>&gt;Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.&lt;/pc<sub>4</sub>&gt;My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.</p>
<p><b>P8 by Arcadian</b> &lt;pa<sub>5</sub>&gt;&lt;a<sub>2</sub>&gt;To simplify the discussion, I have restored that section to your version.&lt;/a<sub>2</sub>&gt;&lt;/pa<sub>5</sub>&gt; You may make any further edits, and &lt;pc<sub>6</sub>&gt;I will have no objection.&lt;/pc<sub>6</sub>&gt;</p>
<p><b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. &lt;pc<sub>7</sub>&gt;It would be defensible to list ALL and CLL here.&lt;/pc<sub>7</sub>&gt;&lt;pa<sub>7</sub>&gt;They fall under the lymphoproliferative disorders.&lt;/pa<sub>7</sub>&gt;</p>

# Agreement and Disagreement

- Examine pairs of phrases to model others' acceptance of the participant's ideas
- Type: Direct or Indirect
- Mode: Response or Paraphrase

**P1 by Arcadian:** There seems to be a much better list at the National Cancer Institute than the one we've got. It ties much better to the actual publication (the same 11 sections, in the same order). I'd like to replace that section in this article. Any objections?

**P2 by JFW:** Not a problem. Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal

Example of Agreement

# Agreement/Disagreement in the Example Thread

<p><b>P1 by Arcadian</b> &lt;pc<sub>1</sub>&gt;There seems to be a much better list at the National Cancer Institute than the one we've got.&lt;/pc<sub>1</sub>&gt;&lt;pa<sub>1</sub>&gt;It ties much better to the actual publication (the same 11 sections, in the same order).&lt;/pa<sub>1</sub>&gt; I'd like to replace that section in this article. Any objections?</p>
<p><b>P2 by JFW</b> &lt;pc<sub>2</sub>&gt;&lt;a<sub>1</sub>&gt;Not a problem.&lt;/a<sub>1</sub>&gt;&lt;/pc<sub>2</sub>&gt;Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal</p>
<p><b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.</p>
<p><b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! &lt;pc<sub>3</sub>&gt;I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histiocytic and mastocytic neoplasms.&lt;/pc<sub>3</sub>&gt;&lt;pa<sub>3</sub>&gt; Remember, Wikipedia is meant to be readable &lt;/pa<sub>3</sub>&gt;by all. Let me know what you think before I delete the non-lymphoma parts.</p>
<p><b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?</p>
<p><b>P6 by Emmanuelm</b> &lt;d<sub>1</sub>&gt;Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.&lt;/d<sub>1</sub>&gt;Would you mind moving the leukemia portion to the leukemia page?</p>
<p><b>P7 by Emmanuelm</b> &lt;pc<sub>4</sub>&gt;Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.&lt;/pc<sub>4</sub>&gt;My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.</p>
<p><b>P8 by Arcadian</b> &lt;pa<sub>5</sub>&gt;&lt;a<sub>2</sub>&gt;To simplify the discussion, I have restored that section to your version.&lt;/a<sub>2</sub>&gt;&lt;/pa<sub>5</sub>&gt;You may make any further edits, and &lt;pc<sub>6</sub>&gt;I will have no objection.&lt;/pc<sub>6</sub>&gt;</p>
<p><b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. &lt;pc<sub>7</sub>&gt;It would be defensible to list ALL and CLL here.&lt;/pc<sub>7</sub>&gt;&lt;pa<sub>7</sub>&gt;They fall under the lymphoproliferative disorders.&lt;/pa<sub>7</sub>&gt;</p>

# Performance of Black Box Components

- The Persuasion and Agreement/Disagreement Language Uses are not described in detail in this work
  - Argumentation is described in (Biran and Rambow, 2011)
  - Claims under submission
  - Agreement/Disagreement annotation is described in (Andreas et al, 2012)

Instead, they are used as black boxes with given performance:

Component	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Attempt to persuade	79.1	69.6	74	57.5	48.2	52.4
Claims	83.6	74.5	78.8	53.7	13.8	22
Argumentation	23.3	91.7	37.1	30.9	48.9	37.8
Agreement	12	31.9	17.4	20	50	28.6
Disagreement	8.7	9.5	9.1	6.3	14.3	8.7

# Dialog Patterns

- Dialog Patterns are based on metadata (e.g., the thread structure), not the text
  - Initiative      who started the thread
  - Investment    share of participation
  - Irrelevance    how often ignored by others
  - Interjection   at what point joined conversation
  - Incitation      how long are branches started



# Dialog Patterns in the Example Thread

- Arcadian has initiative

<b>P1 by Arcadian</b> <i>&lt;pc<sub>1</sub>&gt;</i> There seems to be a much better list at the National Cancer Institute than the one we've got. <i>&lt;/pc<sub>1</sub>&gt;&lt;pa<sub>1</sub>&gt;</i> It ties much better to the actual publication (the same 11 sections, in the same order). <i>&lt;/pa<sub>1</sub>&gt;</i> I'd like to replace that section in this article. Any objections?
<b>P2 by JFW</b> <i>&lt;pc<sub>2</sub>&gt;&lt;a<sub>1</sub>&gt;</i> Not a problem. <i>&lt;/a<sub>1</sub>&gt;&lt;/pc<sub>2</sub>&gt;</i> Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal
<b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.
<b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <i>&lt;pc<sub>3</sub>&gt;</i> I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histocytic and mastocytic neoplasms. <i>&lt;/pc<sub>3</sub>&gt;&lt;pa<sub>3</sub>&gt;</i> Remember, Wikipedia is meant to be readable <i>&lt;/pa<sub>3</sub>&gt;</i> by all. Let me know what you think before I delete the non-lymphoma parts.
<b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?
<b>P6 by Emmanuelm</b> <i>&lt;d<sub>1</sub>&gt;</i> Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th. <i>&lt;/d<sub>1</sub>&gt;</i> Would you mind moving the leukemia portion to the leukemia page?
<b>P7 by Emmanuelm</b> <i>&lt;pc<sub>4</sub>&gt;</i> Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages. <i>&lt;/pc<sub>4</sub>&gt;</i> My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.
<b>P8 by Arcadian</b> <i>&lt;pa<sub>5</sub>&gt;&lt;a<sub>2</sub>&gt;</i> To simplify the discussion, I have restored that section to your version. <i>&lt;/a<sub>2</sub>&gt;&lt;/pa<sub>5</sub>&gt;</i> You may make any further edits, and <i>&lt;pc<sub>6</sub>&gt;</i> I will have no objection. <i>&lt;/pc<sub>6</sub>&gt;</i>
<b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <i>&lt;pc<sub>7</sub>&gt;</i> It would be defensible to list ALL and CLL here. <i>&lt;/pc<sub>7</sub>&gt;&lt;pa<sub>7</sub>&gt;</i> They fall under the lymphoproliferative disorders. <i>&lt;/pa<sub>7</sub>&gt;</i>

# Dialog Patterns in the Example Thread

- Investment: Arcadian 4, JFW 2, Emmanuelm 3

<b>P1 by Arcadian</b> <pc <sub>1</sub> >There seems to be a much better list at the National Cancer Institute than the one we've got.</pc <sub>1</sub> ><pa <sub>1</sub> >It ties much better to the actual publication (the same 11 sections, in the same order).</pa <sub>1</sub> > I'd like to replace that section in this article. Any objections?
<b>P2 by JFW</b> <pc <sub>2</sub> ><a <sub>1</sub> >Not a problem.</a <sub>1</sub> ></pc <sub>2</sub> >Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal
<b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.
<b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <pc <sub>3</sub> >I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histocytic and mastocytic neoplasms.</pc <sub>3</sub> ><pa <sub>3</sub> > Remember, Wikipedia is meant to be readable </pa <sub>3</sub> >by all. Let me know what you think before I delete the non-lymphoma parts.
<b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?
<b>P6 by Emmanuelm</b> <d <sub>1</sub> >Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.</d <sub>1</sub> >Would you mind moving the leukemia portion to the leukemia page?
<b>P7 by Emmanuelm</b> <pc <sub>4</sub> >Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.</pc <sub>4</sub> >My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.
<b>P8 by Arcadian</b> <pa <sub>5</sub> ><a <sub>2</sub> >To simplify the discussion, I have restored that section to your version.</a <sub>2</sub> ></pa <sub>5</sub> >You may make any further edits, and <pc <sub>6</sub> >I will have no objection.</pc <sub>6</sub> >
<b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <pc <sub>7</sub> >It would be defensible to list ALL and CLL here.</pc <sub>7</sub> ><pa <sub>7</sub> >They fall under the lymphoproliferative disorders.</pa <sub>7</sub> >

# Dialog Patterns in the Example Thread

- Irrelevance: **Arcadian 2**, **JFW 1**, **Emmanuelm 1**

<b>P1 by Arcadian</b> <i>&lt;pc<sub>1</sub>&gt;</i> There seems to be a much better list at the National Cancer Institute than the one we've got. <i>&lt;/pc<sub>1</sub>&gt;&lt;pa<sub>1</sub>&gt;</i> It ties much better to the actual publication (the same 11 sections, in the same order). <i>&lt;/pa<sub>1</sub>&gt;</i> I'd like to replace that section in this article. Any objections?
<b>P2 by JFW</b> <i>&lt;pc<sub>2</sub>&gt;&lt;a<sub>1</sub>&gt;</i> Not a problem. <i>&lt;/a<sub>1</sub>&gt;&lt;/pc<sub>2</sub>&gt;</i> Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal
<b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.
<b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <i>&lt;pc<sub>3</sub>&gt;</i> I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histocytic and mastocytic neoplasms. <i>&lt;/pc<sub>3</sub>&gt;&lt;pa<sub>3</sub>&gt;</i> Remember, Wikipedia is meant to be readable <i>&lt;/pa<sub>3</sub>&gt;</i> by all. Let me know what you think before I delete the non-lymphoma parts.
<b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?
<b>P6 by Emmanuelm</b> <i>&lt;d<sub>1</sub>&gt;</i> Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th. <i>&lt;/d<sub>1</sub>&gt;</i> Would you mind moving the leukemia portion to the leukemia page?
<b>P7 by Emmanuelm</b> <i>&lt;pc<sub>4</sub>&gt;</i> Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages. <i>&lt;/pc<sub>4</sub>&gt;</i> My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.
<b>P8 by Arcadian</b> <i>&lt;pa<sub>5</sub>&gt;&lt;a<sub>2</sub>&gt;</i> To simplify the discussion, I have restored that section to your version. <i>&lt;/a<sub>2</sub>&gt;&lt;/pa<sub>5</sub>&gt;</i> You may make any further edits, and <i>&lt;pc<sub>6</sub>&gt;</i> I will have no objection. <i>&lt;/pc<sub>6</sub>&gt;</i>
<b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <i>&lt;pc<sub>7</sub>&gt;</i> It would be defensible to list ALL and CLL here. <i>&lt;/pc<sub>7</sub>&gt;&lt;pa<sub>7</sub>&gt;</i> They fall under the lymphoproliferative disorders. <i>&lt;/pa<sub>7</sub>&gt;</i>

# Dialog Patterns in the Example Thread

- Interjection: **Arcadian 1**, **JFW 2**, **Emmanuelm 4**

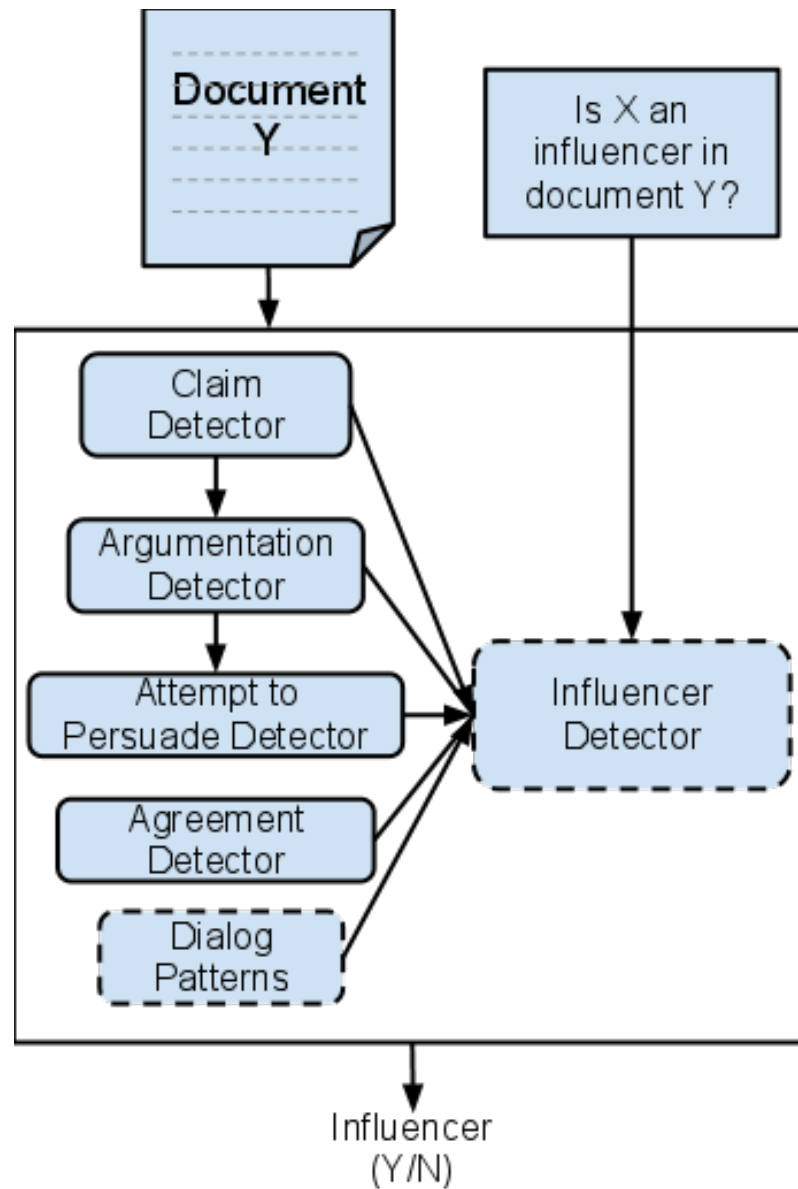
<b>P1 by Arcadian</b> <pc <sub>1</sub> >There seems to be a much better list at the National Cancer Institute than the one we've got.</pc <sub>1</sub> ><pa <sub>1</sub> >It ties much better to the actual publication (the same 11 sections, in the same order).</pa <sub>1</sub> > I'd like to replace that section in this article. Any objections?
<b>P2 by JFW</b> <pc <sub>2</sub> ><a <sub>1</sub> >Not a problem.</a <sub>1</sub> ></pc <sub>2</sub> >Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal
<b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.
<b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <pc <sub>3</sub> >I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histocytic and mastocytic neoplasms.</pc <sub>3</sub> ><pa <sub>3</sub> > Remember, Wikipedia is meant to be readable </pa <sub>3</sub> >by all. Let me know what you think before I delete the non-lymphoma parts.
<b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?
<b>P6 by Emmanuelm</b> <d <sub>1</sub> >Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.</d <sub>1</sub> >Would you mind moving the leukemia portion to the leukemia page?
<b>P7 by Emmanuelm</b> <pc <sub>4</sub> >Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.</pc <sub>4</sub> >My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.
<b>P8 by Arcadian</b> <pa <sub>5</sub> ><a <sub>2</sub> >To simplify the discussion, I have restored that section to your version.</a <sub>2</sub> ></pa <sub>5</sub> >You may make any further edits, and <pc <sub>6</sub> >I will have no objection.</pc <sub>6</sub> >
<b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <pc <sub>7</sub> >It would be defensible to list ALL and CLL here.</pc <sub>7</sub> ><pa <sub>7</sub> >They fall under the lymphoproliferative disorders.</pa <sub>7</sub> >

# Dialog Patterns in the Example Thread

- Incitation: **Arcadian 4**, **JFW 1**, **Emmanuelm 3**

<b>P1 by Arcadian</b> <pc <sub>1</sub> >There seems to be a much better list at the National Cancer Institute than the one we've got.</pc <sub>1</sub> ><pa <sub>1</sub> >It ties much better to the actual publication (the same 11 sections, in the same order).</pa <sub>1</sub> > I'd like to replace that section in this article. Any objections?
<b>P2 by JFW</b> <pc <sub>2</sub> ><a <sub>1</sub> >Not a problem.</a <sub>1</sub> ></pc <sub>2</sub> >Perhaps we can also insert the relative incidence as published in this month's wiki Blood journal
<b>P3 by Arcadian</b> I've made the update. I've included template links to a source that supports looking up information by ICD-O code.
<b>P4 by Emmanuelm</b> Can Arcadian tell me why he/she included the leukemia classification to this lymphoma page? It is not even listed in the Wikipedia leukemia page! <pc <sub>3</sub> >I vote for dividing the WHO classification into 4 parts in 4 distinct pages: leukemia, lymphoma, histocytic and mastocytic neoplasms.</pc <sub>3</sub> ><pa <sub>3</sub> > Remember, Wikipedia is meant to be readable </pa <sub>3</sub> >by all. Let me know what you think before I delete the non-lymphoma parts.
<b>P5 by Arcadian</b> Emmanuelm, aren't you the person who added those other categories on 6 July 2005?
<b>P6 by Emmanuelm</b> <d <sub>1</sub> >Arcadian, I added only the lymphoma portion of the WHO classification. You added the leukemias on Dec 29th.</d <sub>1</sub> >Would you mind moving the leukemia portion to the leukemia page?
<b>P7 by Emmanuelm</b> <pc <sub>4</sub> >Oh, and please note that I would be very comfortable with a "cross-coverage" of lymphocytic leukemias in both pages.</pc <sub>4</sub> >My comment is really about myeloid, histiocytic and mast cell neoplasms who share no real relationship with lymphomas.
<b>P8 by Arcadian</b> <pa <sub>5</sub> ><a <sub>2</sub> >To simplify the discussion, I have restored that section to your version.</a <sub>2</sub> ></pa <sub>5</sub> >You may make any further edits, and <pc <sub>6</sub> >I will have no objection.</pc <sub>6</sub> >
<b>P9 by JFW</b> The full list should be on the hematological malignancy page, and the lymphoma part can be here. <pc <sub>7</sub> >It would be defensible to list ALL and CLL here.</pc <sub>7</sub> ><pa <sub>7</sub> >They fall under the lymphoproliferative disorders.</pa <sub>7</sub> >

# System Pipeline



# Experiments

- “Is person X an influencer in thread Y?”
- Separate experiments for Wikipedia Discussions and LiveJournal
- Only threads in which there was at least 1 influencer
  
- SVM using the Language Uses as features
- 10-fold cross validation
- Tried each Language Use separately as well as permutations

# Results

- Baseline all-positive (everyone is an influencer)

System	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Baseline: all-yes	16.2	100	27.9	19.2	19.2	32.2
Full	40.5	80.5	53.9	61.7	82	70.4
Initiative	31.6	31.2	31.4	73.5	72.7	73.1
Irrelevance	21.7	77.9	34	19.2	100	32.2
Incitation	28.3	77.9	41.5	49.5	73.8	59.2
Investment	43	71.4	53.7	50.2	75.4	60.3
Interjection	24.7	88.3	38.6	36.9	91.3	52.5
Agreement	36	46.8	40.7	45.1	82.5	58.3
Disagreement	35.3	70.1	47	19.2	100	32.2
Claims	40	72.7	51.6	54.3	76	63.3
Argumentation	19	98.7	31.8	31.1	85.2	45.6
Attempt to persuade	23.7	79.2	36.5	37.4	48.1	42.1
<b>Best system</b>	<b>47</b>	<b>80.5</b>	<b>59.3</b>	<b>66.2</b>	<b>84.7</b>	<b>74.3</b>



# Results

- Full systems make use of all 10 Language Uses

System	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Baseline: all-yes	16.2	100	27.9	19.2	19.2	32.2
<b>Full</b>	<b>40.5</b>	<b>80.5</b>	<b>53.9</b>	<b>61.7</b>	<b>82</b>	<b>70.4</b>
Initiative	31.6	31.2	31.4	73.5	72.7	73.1
Irrelevance	21.7	77.9	34	19.2	100	32.2
Incitation	28.3	77.9	41.5	49.5	73.8	59.2
Investment	43	71.4	53.7	50.2	75.4	60.3
Interjection	24.7	88.3	38.6	36.9	91.3	52.5
Agreement	36	46.8	40.7	45.1	82.5	58.3
Disagreement	35.3	70.1	47	19.2	100	32.2
Claims	40	72.7	51.6	54.3	76	63.3
Argumentation	19	98.7	31.8	31.1	85.2	45.6
Attempt to persuade	23.7	79.2	36.5	37.4	48.1	42.1
<b>Best system</b>	<b>47</b>	<b>80.5</b>	<b>59.3</b>	<b>66.2</b>	<b>84.7</b>	<b>74.3</b>

# Results

- We tried each component individually, before looking at all permutations

System	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Baseline: all-yes	16.2	100	27.9	19.2	19.2	32.2
Full	40.5	80.5	53.9	61.7	82	70.4
Initiative	31.6	31.2	31.4	73.5	72.7	73.1
Irrelevance	21.7	77.9	34	19.2	100	32.2
Incitation	28.3	77.9	41.5	49.5	73.8	59.2
Investment	43	71.4	53.7	50.2	75.4	60.3
Interjection	24.7	88.3	38.6	36.9	91.3	52.5
Agreement	36	46.8	40.7	45.1	82.5	58.3
Disagreement	35.3	70.1	47	19.2	100	32.2
Claims	40	72.7	51.6	54.3	76	63.3
Argumentation	19	98.7	31.8	31.1	85.2	45.6
Attempt to persuade	23.7	79.2	36.5	37.4	48.1	42.1
Best system	<b>47</b>	<b>80.5</b>	<b>59.3</b>	<b>66.2</b>	<b>84.7</b>	<b>74.3</b>

# Results

- We tried each component individually, before looking at all permutations

System	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Baseline: all-yes	16.2	100	27.9	19.2	19.2	32.2
Full	40.5	80.5	53.9	61.7	82	70.4
Initiative	31.6	31.2	31.4	73.5	72.7	73.1
Irrelevance	21.7	77.9	34	19.2	100	32.2
Incitation	28.3	77.9	41.5	49.5	73.8	59.2
Investment	43	71.4	53.7	50.2	75.4	60.3
Interjection	24.7	88.3	38.6	36.9	91.3	52.5
Agreement	36	46.8	40.7	45.1	82.5	58.3
Disagreement	35.3	70.1	47	19.2	100	32.2
Claims	40	72.7	51.6	54.3	76	63.3
Argumentation	19	98.7	31.8	31.1	85.2	45.6
Attempt to persuade	23.7	79.2	36.5	37.4	48.1	42.1
Best system	<b>47</b>	<b>80.5</b>	<b>59.3</b>	<b>66.2</b>	<b>84.7</b>	<b>74.3</b>

# Results

- Best system different for each corpus

- Wikipedia:

- Claims
- Argumentation
- Agreement
- Investment

- LiveJournal:

- Attempt to Persuade
- Argumentation
- all 5 Dialog Patterns

System	Wikipedia			LiveJournal		
	P	R	F	P	R	F
Baseline: all-yes	16.2	100	27.9	19.2	19.2	32.2
Full	40.5	80.5	53.9	61.7	82	70.4
Initiative	31.6	31.2	31.4	73.5	72.7	73.1
Irrelevance	21.7	77.9	34	19.2	100	32.2
Incitation	28.3	77.9	41.5	49.5	73.8	59.2
Investment	43	71.4	53.7	50.2	75.4	60.3
Interjection	24.7	88.3	38.6	36.9	91.3	52.5
Agreement	36	46.8	40.7	45.1	82.5	58.3
Disagreement	35.3	70.1	47	19.2	100	32.2
Claims	40	72.7	51.6	54.3	76	63.3
Argumentation	19	98.7	31.8	31.1	85.2	45.6
Attempt to persuade	23.7	79.2	36.5	37.4	48.1	42.1
<b>Best system</b>	<b>47</b>	<b>80.5</b>	<b>59.3</b>	<b>66.2</b>	<b>84.7</b>	<b>74.3</b>

# Perfect Components Evaluation

- Question: do errors come from bad choice of component, or from imperfection of component?
  - We can check using the gold data instead of the component
  - To avoid noise from other components, look at each component alone (make system detect influencer based solely on component output)

- Example: attempt to persuade

Data Set	Our System			Gold Answers		
	P	R	F	P	R	F
Wikipedia	23.6	69.4	35.2	23.8	81.6	36.9
LiveJournal	37.5	48.1	42.1	40.7	61.8	49

- Recall that the Attempt to Persuade black box has internal performance of 74 (F-Measure) in Wikipedia and 52.4 in LiveJournal

# Wikipedia Discussions VS LiveJournal

LiveJournal	Wikipedia Discussions
Easier	Harder
Initiative is very strong	
All dialog patterns are useful	Only investment is useful
Attempt to persuade is useful	Agreement is useful

# Conclusion

- Detecting influencers is hard
- F-measure is 74.3 in LiveJournal and 59.3 in Wikipedia Discussions
  - LiveJournal is easier
  - Both significantly exceed the baselines
- Using perfect components improves performance
- The usefulness of Language Uses differs depending on the dataset

# Questions?