

# Predicting Prerequisite Structure in Wikipedia



Partha Pratim Talukdar  
(CMU)



William W. Cohen  
(CMU)

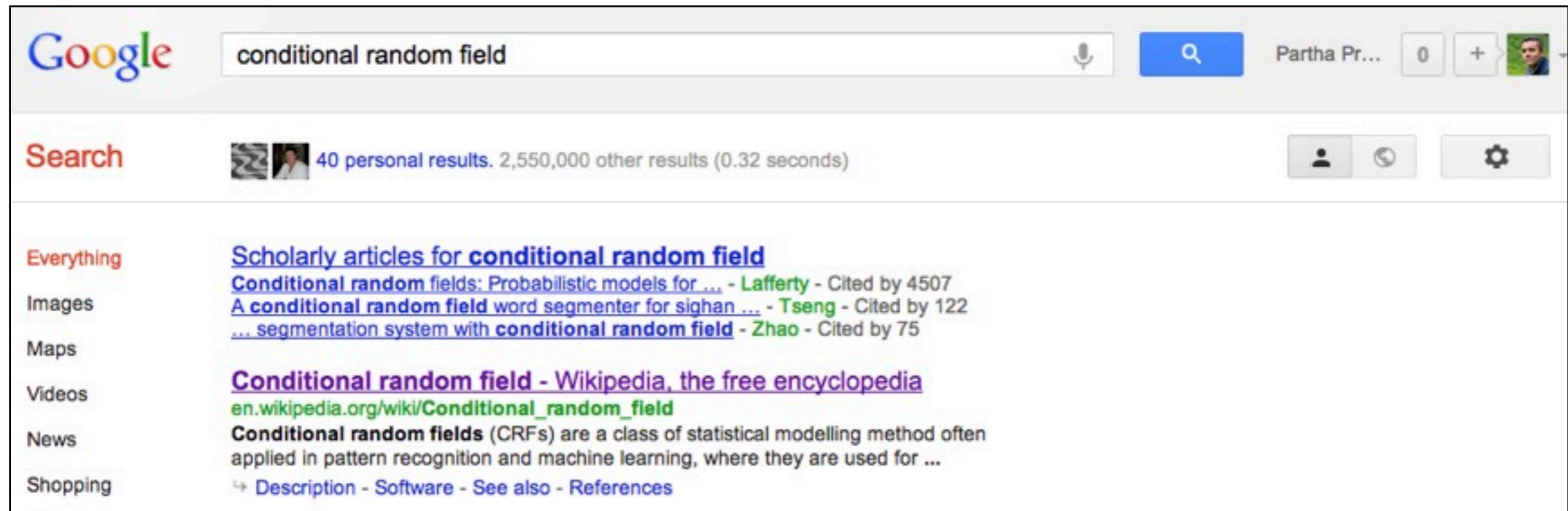
*7<sup>th</sup> Workshop on Innovative Use of NLP for Building Educational Applications (BEA7)  
NAACL-HLT 2012, June 7, 2012*

Need to Comprehend: *Conditional Random Field*

# Need to Comprehend: *Conditional Random Field*

The image shows a Google search interface. At the top, the Google logo is on the left, and the search bar contains the text "conditional random field". To the right of the search bar are a microphone icon, a blue search button with a magnifying glass, and a user profile icon labeled "Partha Pr...". Below the search bar, the word "Search" is displayed in red. To its right, there are icons for a person, a refresh button, and a settings gear. The main content area shows search results. On the left side, there is a vertical menu with options: "Everything", "Images", "Maps", "Videos", "News", "Shopping", and "More". Below this menu, there is a section for "Pittsburgh, PA" with a "Change location" link and a "Show search tools" link. The search results are listed as follows:  
1. **Scholarly articles for conditional random field**  
- [Conditional random fields: Probabilistic models for ...](#) - Lafferty - Cited by 4507  
- [A conditional random field word segmenter for sghan ...](#) - Tseng - Cited by 122  
- [... segmentation system with conditional random field](#) - Zhao - Cited by 75  
2. **Conditional random field - Wikipedia, the free encyclopedia**  
- [en.wikipedia.org/wiki/Conditional\\_random\\_field](http://en.wikipedia.org/wiki/Conditional_random_field)  
- **Conditional random fields** (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for ...  
- [↪ Description](#) - [Software](#) - [See also](#) - [References](#)  
3. **[PDF] Conditional Random Fields: An Introduction**  
- [www.inference.phy.cam.ac.uk/hmw26/papers/crf\\_intro.pdf](http://www.inference.phy.cam.ac.uk/hmw26/papers/crf_intro.pdf)  
- File Format: PDF/Adobe Acrobat - [Quick View](#)  
- by HM Wallach - 2004 - [Cited by 169](#) - [Related articles](#)  
- **Conditional Random Fields: An Introduction.** \*. Hanna M. Wallach. February 24, 2004.  
- 1 Labeling Sequential Data. The task of assigning label sequences to a ...  
4. **Conditional Random Fields**  
- [www.inference.phy.cam.ac.uk/hmw26/crf/](http://www.inference.phy.cam.ac.uk/hmw26/crf/)  
- This page contains material on, or relating to, **conditional random fields**. I shall continue to update this page as research on **conditional random fields** advances, ...  
5. **[PDF] Conditional Random Fields: Probabilistic Models for Segmenti...**  
- [www.cis.upenn.edu/~pereira/papers/crf.pdf](http://www.cis.upenn.edu/~pereira/papers/crf.pdf)  
- File Format: PDF/Adobe Acrobat - [Quick View](#)  
- by J Lafferty - [Cited by 4495](#) - [Related articles](#)  
- This paper introduces **conditional random fields** (CRFs), a sequence modeling ... each other. We can also think of a **CRF** as a finite state model with un- ...

# Need to Comprehend: *Conditional Random Field*



The screenshot shows a Google search interface. The search bar contains the text "conditional random field". The search results are displayed under the "Everything" tab. The first result is a link to "Scholarly articles for conditional random field". Below this, there are three entries for scholarly articles: "Conditional random fields: Probabilistic models for ..." by Lafferty (4507 citations), "A conditional random field word segmenter for sghan ..." by Tseng (122 citations), and "... segmentation system with conditional random field" by Zhao (75 citations). The second result is the Wikipedia page for "Conditional random field", with a snippet: "Conditional random fields (CRFs) are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for ...". The left sidebar shows navigation options: Search, Images, Maps, Videos, News, and Shopping.

What to read first and how to go about?

This page contains material on, or relating to, **conditional random fields**. I shall continue to update this page as research on **conditional random fields** advances, ...

[PDF] [Conditional Random Fields: Probabilistic Models for Segmenti...](#)

[www.cis.upenn.edu/~pereira/papers/crf.pdf](http://www.cis.upenn.edu/~pereira/papers/crf.pdf)

File Format: PDF/Adobe Acrobat - Quick View

by J Lafferty - Cited by 4495 - Related articles

This paper introduces **conditional random fields** (CRFs), a sequence modeling ... each other. We can also think of a **CRF** as a finite state model with un- ...



WIKIPEDIA  
The Free Encyclopedia

[Main page](#)  
[Contents](#)  
[Featured content](#)  
[Current events](#)  
[Random article](#)  
[Donate to Wikipedia](#)

▼ [Interaction](#)  
[Help](#)  
[About Wikipedia](#)  
[Community portal](#)  
[Recent changes](#)  
[Contact Wikipedia](#)

► [Toolbox](#)  
► [Print/export](#)

▼ [Languages](#)  
[Deutsch](#)  
[Español](#)  
[日本語](#)  
[中文](#)

[Log in / create account](#)

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

## Conditional random field

From Wikipedia, the free encyclopedia

This article **provides insufficient context for those unfamiliar with the subject**. Please help [improve the article](#) with a [good introductory style](#). *(June 2011)*

**Conditional random fields (CRFs)** are a class of [statistical modelling method](#) often applied in [pattern recognition](#) and [machine learning](#), where they are used for [structured prediction](#). Whereas an ordinary [classifier](#) predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in [natural language processing](#) predicts sequences of labels for sequences of input samples.

CRFs are a type of [discriminative undirected probabilistic graphical model](#). It is used to encode known relationships between observations and construct consistent interpretations. It is often used for [labeling](#) or [parsing](#) of sequential data, such as natural language text or [biological sequences](#)<sup>[1]</sup> and in [computer vision](#).<sup>[2]</sup> Specifically, CRFs find applications in [shallow parsing](#),<sup>[3]</sup> [named entity recognition](#)<sup>[4]</sup> and [gene finding](#), among other tasks, being an alternative to the related [hidden Markov models](#). In computer vision, CRFs are often used for object recognition and image segmentation.

### Contents [\[hide\]](#)

- 1 Description
  - 1.1 Inference
  - 1.2 Parameter Learning
  - 1.3 Examples
  - 1.4 Higher-order CRFs and semi-Markov CRFs
- 2 Software
- 3 See also
- 4 References

[Visit the main page](#)



WIKIPEDIA  
The Free Encyclopedia

- [Main page](#)
- [Contents](#)
- [Featured content](#)
- [Current events](#)
- [Random article](#)
- [Donate to Wikipedia](#)

- Interaction
  - [Help](#)
  - [About Wikipedia](#)
  - [Community portal](#)
  - [Recent changes](#)
  - [Contact Wikipedia](#)


- [Toolbox](#)
- [Print/export](#)

- Languages
  - [Deutsch](#)
  - [Español](#)
  - [日本語](#)
  - [中文](#)

Article [Talk](#) [Read](#)

# Conditional random field

From Wikipedia, the free encyclopedia

 This article **provides insufficient coverage** on this **subject**. Please help [improve the article](#) (see the talk page for details). Please do not remove this message until the problem has been solved. (Help) (2011)

**Conditional random fields (CRFs)** are a class of **statistical modelling method** often applied in **pattern recognition** and **machine learning**, where they are used for **structured prediction**. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in **natural language processing** predicts sequences of labels for sequences of input samples.

CRFs are a type of **discriminative undirected probabilistic graphical model**. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for **labeling** or **parsing** of sequential data, such as natural language text or **biological sequences**<sup>[1]</sup> and in **computer vision**.<sup>[2]</sup> Specifically, CRFs find applications in **shallow parsing**,<sup>[3]</sup> **named entity recognition**<sup>[4]</sup> and **gene finding**, among other tasks, being an alternative to the related **hidden Markov models**. In computer vision, CRFs are often used for object recognition and image segmentation.

**Contents** [\[hide\]](#)

- 1 Description
  - 1.1 Inference
  - 1.2 Parameter Learning
  - 1.3 Examples
  - 1.4 Higher-order CRFs and semi-Markov CRFs
- 2 Software
- 3 See also
- 4 References

[Visit the main page](#)

WIKIPEDIA  
The Free Encyclopedia

Article [Talk](#) [Read](#) [Edit](#) [View history](#)

## Graphical model

From Wikipedia, the free encyclopedia

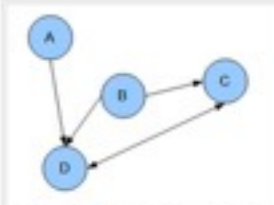
A **graphical model** is a probabilistic model for which a graph denotes the conditional independence structure between random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

**Contents** [\[hide\]](#)

- 1 Types of graphical models
  - 1.1 Bayesian network
  - 1.2 Markov random field
  - 1.3 Other types
- 2 Applications
- 3 See also
- 4 Notes
- 5 References and further reading
  - 5.1 Books and book chapters
  - 5.2 Journal articles
  - 5.3 Other

**Types of graphical models** [\[edit\]](#)

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or *factorized* representation of a set of independences that hold in the specific distribution. Two branches of graphical representations of distributions are commonly used, namely, *Bayesian networks* and *Markov networks*. Both families encompass the properties of factorization and independences, but they differ in the set of independences they can encode and the factorization of



An example of a graphical model. Each arrow indicates a dependency. In this example: D depends on A, D depends on B, D depends on C, C depends on B, and C depends on D.

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Random variable

From Wikipedia, the free encyclopedia

In probability and statistics, a **random variable** or **stochastic variable** is a **variable** whose value is subject to variations due to chance (i.e. *randomness*, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated *probability*. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: as the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a *die*, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like *Bill Gates* after Internet research on the subject, which might have possible values ranging between about \$50 billion and \$100 billion, with values near the center more likely).

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Graphical model

From Wikipedia, the free encyclopedia

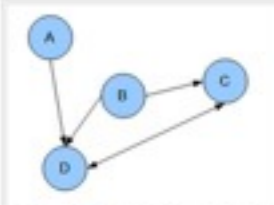
A **graphical model** is a probabilistic model for which a graph denotes the conditional independence structure of the model's random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

**Contents** [hide]

- Types of graphical models
  - Bayesian network
  - Markov random field
  - Other types
- Applications
- See also
- Notes
- References and further reading
  - Books and book chapters
  - Journal articles
  - Other

**Types of graphical models** [edit]

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or *factorized* representation of a set of independences that hold in the specific distribution. Two branches of graphical representations of distributions are commonly used, namely, *Bayesian networks* and *Markov networks*. Both families encompass the properties of factorization and independences, but they differ in the set of independences they can encode and the factorization of



An example of a graphical model. Each arrow indicates a dependency. In this example: D depends on A, D depends on B, D depends on C, C depends on B, and C depends on D.

Current events

Random article

Donate to Wikipedia

Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia

Toolbox

Print/export

Languages

- Deutsch
- Español
- 日本語
- 中文

**Conditional random fields (CRFs)** are a class of **statistical modelling method** often applied in **pattern recognition** and **machine learning**, where they are used for **structured prediction**. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in **natural language processing** predicts sequences of labels for sequences of input samples.

CRFs are a type of **discriminative undirected probabilistic graphical model**. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for **labeling** or **parsing** of sequential data, such as natural language text or **biological sequences**<sup>[1]</sup> and in **computer vision**.<sup>[2]</sup> Specifically, CRFs find applications in **shallow parsing**,<sup>[3]</sup> **named entity recognition**<sup>[4]</sup> and **gene finding**, among other tasks, being an alternative to the related **hidden Markov models**. In computer vision, CRFs are often used for object recognition and image segmentation.

**Contents** [hide]

- Description
  - Inference
  - Parameter Learning
  - Examples
  - Higher-order CRFs and semi-Markov CRFs
- Software
- See also
- References

Visit the main page

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Random variable

From Wikipedia, the free encyclopedia

In probability and statistics, a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: as the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a die, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like Bill Gates after Internet research on the subject, which might have possible values ranging between about \$50 billion and \$100 billion, with values near the center more likely).

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Graphical model

From Wikipedia, the free encyclopedia

A graphical model is a probabilistic model for which a graph denotes the conditional independence structure of the model's random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

Contents [hide]

- Types of graphical models
  - Bayesian network
  - Markov random field
  - Other types
- Applications
- See also
- Notes
- References and further reading
  - Books and book chapters
  - Journal articles
  - Other

Types of graphical models [edit]

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or factorized representation of a set of independences that hold in the specific distribution. Two branches of graphical representations of distributions are commonly used, namely, Bayesian networks and Markov networks. Both families encompass the properties of factorization and independences, but they differ in the types of independences they can encode and the factorization of

An example of a graphical model. Each arrow indicates a dependency. In this example: D depends on A, D depends on B, D depends on C, C depends on B, and C depends on D.

Current events

Random article

Donate to Wikipedia

Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact Wikipedia

Toolbox

Print/export

Languages

- Deutsch
- Español
- 日本語
- 中文

**Conditional random fields (CRFs)** are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in natural language processing predicts sequences of labels for sequences of input samples.

CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences<sup>[1]</sup> and in computer vision.<sup>[2]</sup> Specifically, CRFs find applications in shallow parsing,<sup>[3]</sup> named entity recognition<sup>[4]</sup> and pose finding, among other tasks, helping to

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Markov random field

From Wikipedia, the free encyclopedia

(Redirected from Markov network)

In the domain of physics and probability, a **Markov random field** (often abbreviated as MRF), **Markov network** or **undirected graphical model** is a set of random variables having a Markov property described by an undirected graph. A Markov random field is similar to a Bayesian network in its representation of dependencies; the differences being that Bayesian networks are directed and acyclic, whereas Markov networks are undirected and may be cyclic. Thus, a Markov network can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies); on the other hand, it can't represent certain dependencies that a Bayesian network can (such as induced dependencies).



WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Random variable

From Wikipedia, the free encyclopedia

In probability and statistics, a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: as the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a die, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like Bill Gates after Internet research on the subject, which might have possible values ranging between about \$50 billion and \$100 billion, with values near the center more likely).

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Graphical model

From Wikipedia, the free encyclopedia

A graphical model is a probabilistic model for which a graph denotes the conditional independence structure of the model's random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

Contents [hide]

- Types of graphical models
  - Bayesian network
  - Markov random field
  - Other types
- Applications
- See also
- Notes
- References and further reading
  - Books and book chapters
  - Journal articles
  - Other

Types of graphical models [edit]

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or factorized representation of a set of independences that hold in the specific distribution. Two branches of graphical representations of distributions are commonly used, namely, Bayesian networks and Markov networks. Both families encompass the properties of factorization and independences, but they differ in the types of independences they can encode and the factorization of

An example of a graphical model. Each arrow indicates a dependency. In this example: D depends on A, D depends on B, D depends on C, C depends on B, and C depends on D.

Current events

Random article

Donate to Wikipedia

Interaction

Help

About Wikipedia

Community portal

Recent changes

Contact Wikipedia

Toolbox

Print/export

Languages

Deutsch

Español

日本語

中文

**Conditional random fields (CRFs)** are a class of statistical modelling method often applied in pattern recognition and machine learning, where they are used for structured prediction. Whereas an ordinary classifier predicts a label for a single sample without regard to "neighboring" samples, a CRF can take context into account; e.g., the linear chain CRF popular in natural language processing predicts sequences of labels for sequences of input samples. CRFs are a type of discriminative undirected probabilistic graphical model. It is used to encode known relationships between observations and construct consistent interpretations. It is often used for labeling or parsing of sequential data, such as natural language text or biological sequences<sup>[1]</sup> and in computer vision.<sup>[2]</sup> Specifically, CRFs find applications in shallow parsing,<sup>[3]</sup> named entity recognition<sup>[4]</sup> and pose finding, among other tasks, helping

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Markov random field

From Wikipedia, the free encyclopedia

(Redirected from Markov network)

In the domain of physics and probability, a **Markov random field** (often abbreviated as MRF), **Markov network** or **undirected graphical model** is a set of random variables having Markov property described by an undirected graph. A Markov random field is similar to a Bayesian network in its representation of dependencies; the differences being that Bayesian networks are directed and acyclic, whereas Markov networks are undirected and may be cyclic. Thus, a Markov network can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies); on the other hand, it can't represent certain dependencies that a Bayesian network can (such as induced dependencies).

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Random variable

From Wikipedia, the free encyclopedia

In probability and statistics, a **random variable** or **stochastic variable** is a variable whose value is subject to variations due to chance (i.e. randomness, in a mathematical sense). As opposed to other mathematical variables, a random variable conceptually does not have a single, fixed value (even if unknown); rather, it can take on a set of possible different values, each with an associated probability. The interpretation of a random variable depends on the interpretation of probability:

- The objectivist viewpoint: as the outcome of an experiment or event where randomness is involved (e.g. the result of rolling a die, which is a number between 1 and 6, all with equal probability; or the sum of the results of rolling two dice, which is a number between 2 and 12, with some numbers more likely than others).
- The subjectivist viewpoint: the formal encoding of one's beliefs about the various potential values of a quantity that is not known with certainty (e.g. a particular person's belief about the net worth of someone like Bill Gates).

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Graphical model

From Wikipedia, the free encyclopedia

A graphical model is a probabilistic model for which a graph denotes the conditional independence structure of the model's random variables. They are commonly used in probability theory, statistics—particularly Bayesian statistics—and machine learning.

**Contents** [hide]

- Types of graphical models
  - Bayesian network
  - Markov random field
  - Other types
- Applications
- See also
- Notes
- References and further reading
  - Books and book chapters
  - Journal articles
  - Other

**Types of graphical models** [edit]

Generally, probabilistic graphical models use a graph-based representation as the foundation for encoding a complete distribution over a multi-dimensional space and a graph that is a compact or factorized representation of a complete distribution over a multi-dimensional space.

An example of a graphical model. Each arrow indicates a dependency. In this example: D depends on A, D depends on B, D depends on C, C depends on B, and C depends on D.

Attempting to comprehend a topic top-down can easily become difficult to manage: need better tools!

Community portal

Recent changes

Contact Wikipedia

Toolbox

Print/export

Languages

- Deutsch
- Español
- 日本語
- 中文

WIKIPEDIA The Free Encyclopedia

Log in / create account

Article Talk

Read Edit View history Search

## Markov random field

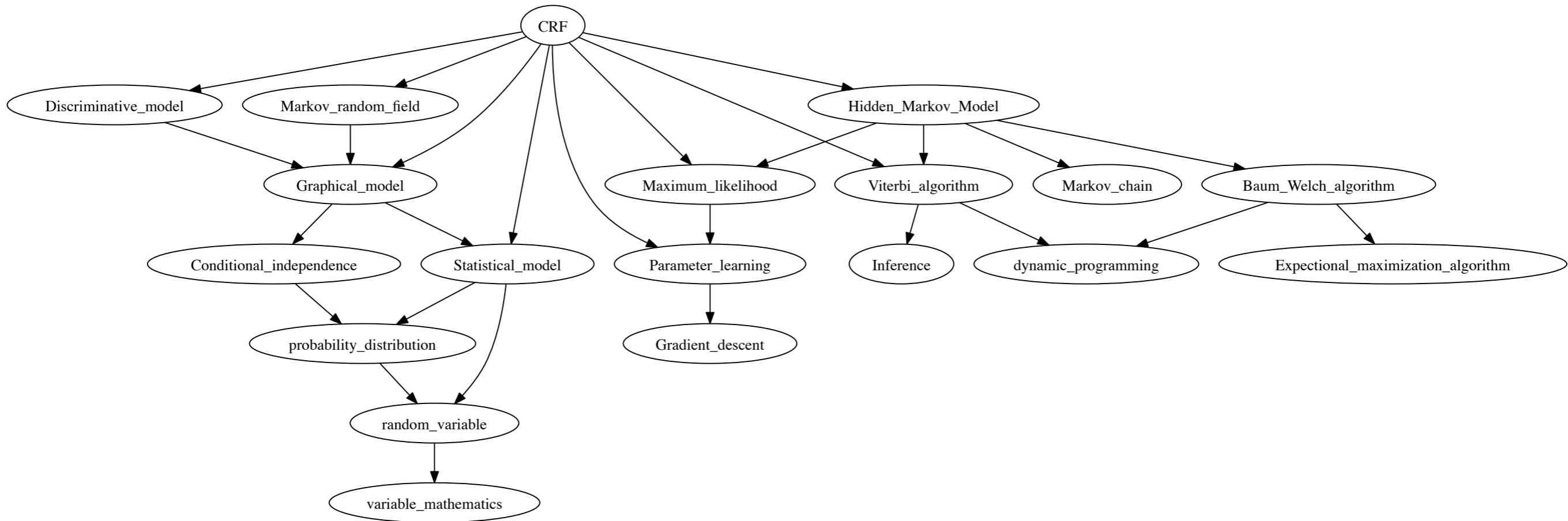
From Wikipedia, the free encyclopedia

(Redirected from Markov network)

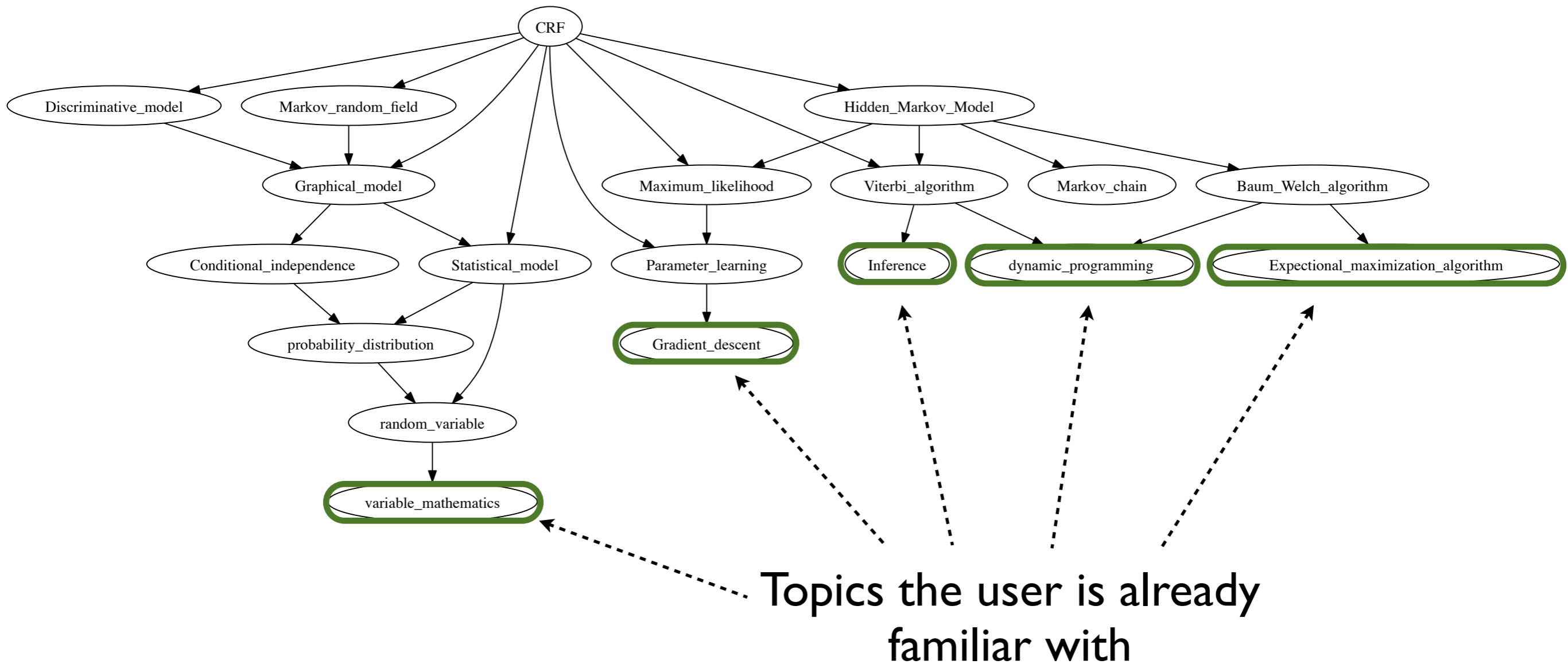
In the domain of physics and probability, a **Markov random field** (often abbreviated as MRF), **Markov network** or **undirected graphical model** is a set of random variables having **Markov property** described by an undirected graph. A Markov random field is similar to a Bayesian network in its representation of dependencies; the differences being that Bayesian networks are directed and acyclic, whereas Markov networks are undirected and may be cyclic. Thus, a Markov network can represent certain dependencies that a Bayesian network cannot (such as cyclic dependencies); on the other hand, it can't represent certain dependencies that a Bayesian network can (such as induced dependencies).

# Comprehension Plan for Conditional Random Field (CRF)

# Comprehension Plan for Conditional Random Field (CRF)



# Comprehension Plan for Conditional Random Field (CRF)

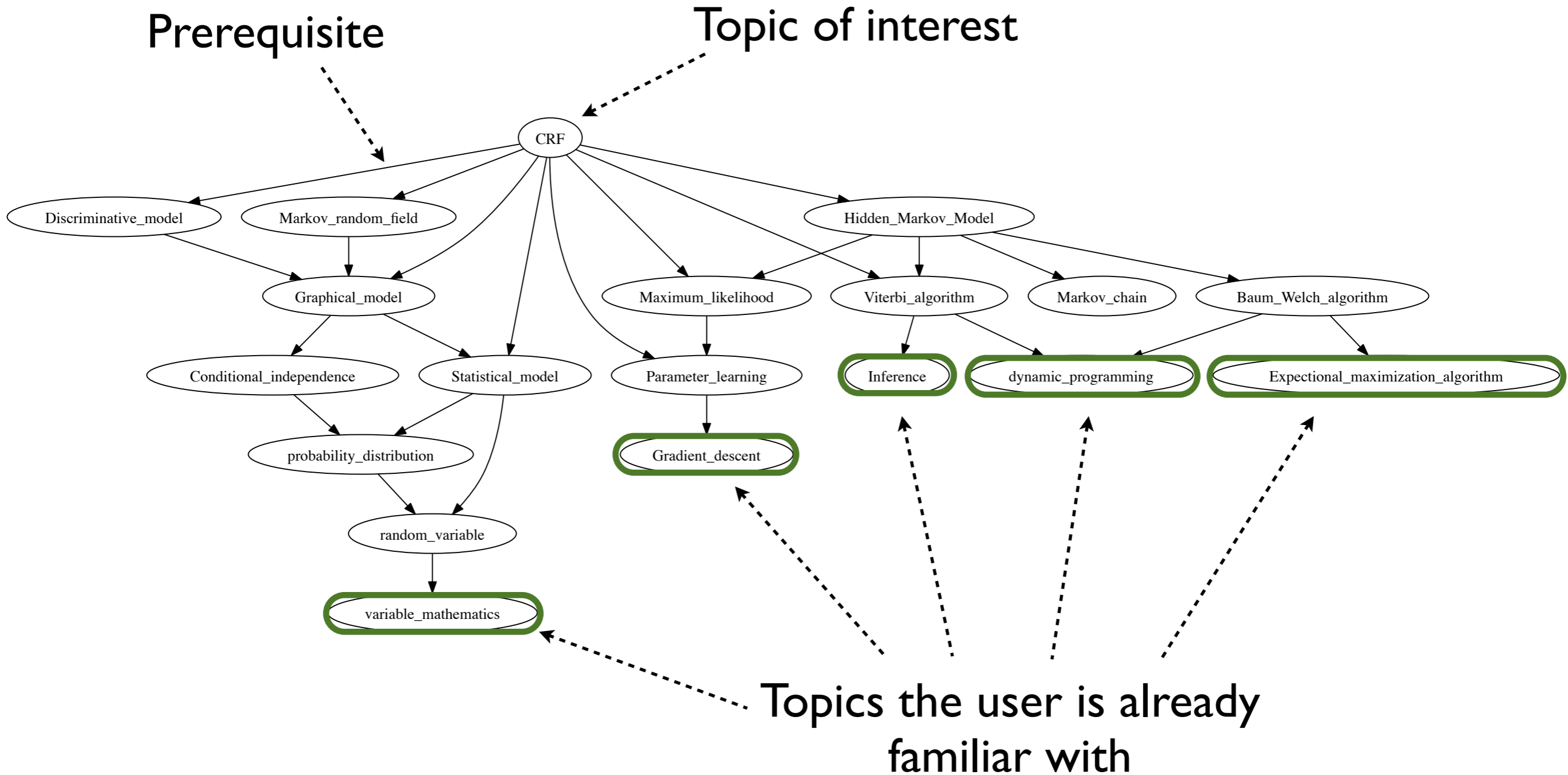




# Comprehension Plan for Conditional Random Field (CRF)

Prerequisite

Topic of interest



# Ingredients for a Comprehension Plan



# Ingredients for a Comprehension Plan

- Estimation of user's current knowledge

# Ingredients for a Comprehension Plan

- Estimation of user's current knowledge
- Determining prerequisite structure among topics

# Ingredients for a Comprehension Plan

- Estimation of user's current knowledge
- Determining prerequisite structure among topics
- Plan over prerequisite structure to connect user's current knowledge with target topic

# Ingredients for a Comprehension Plan

- Estimation of user's current knowledge

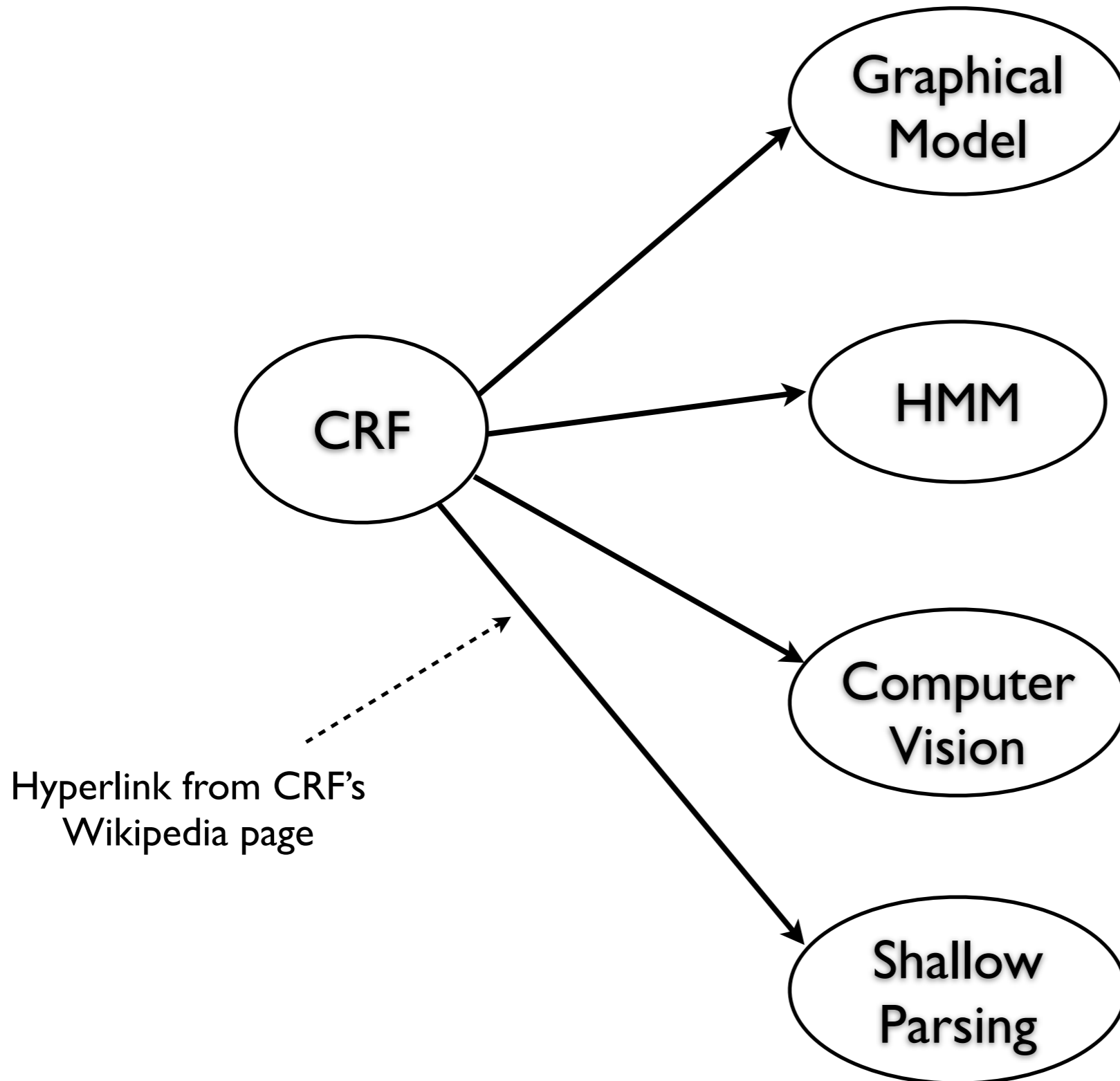
This Talk

Determining prerequisite structure among topics

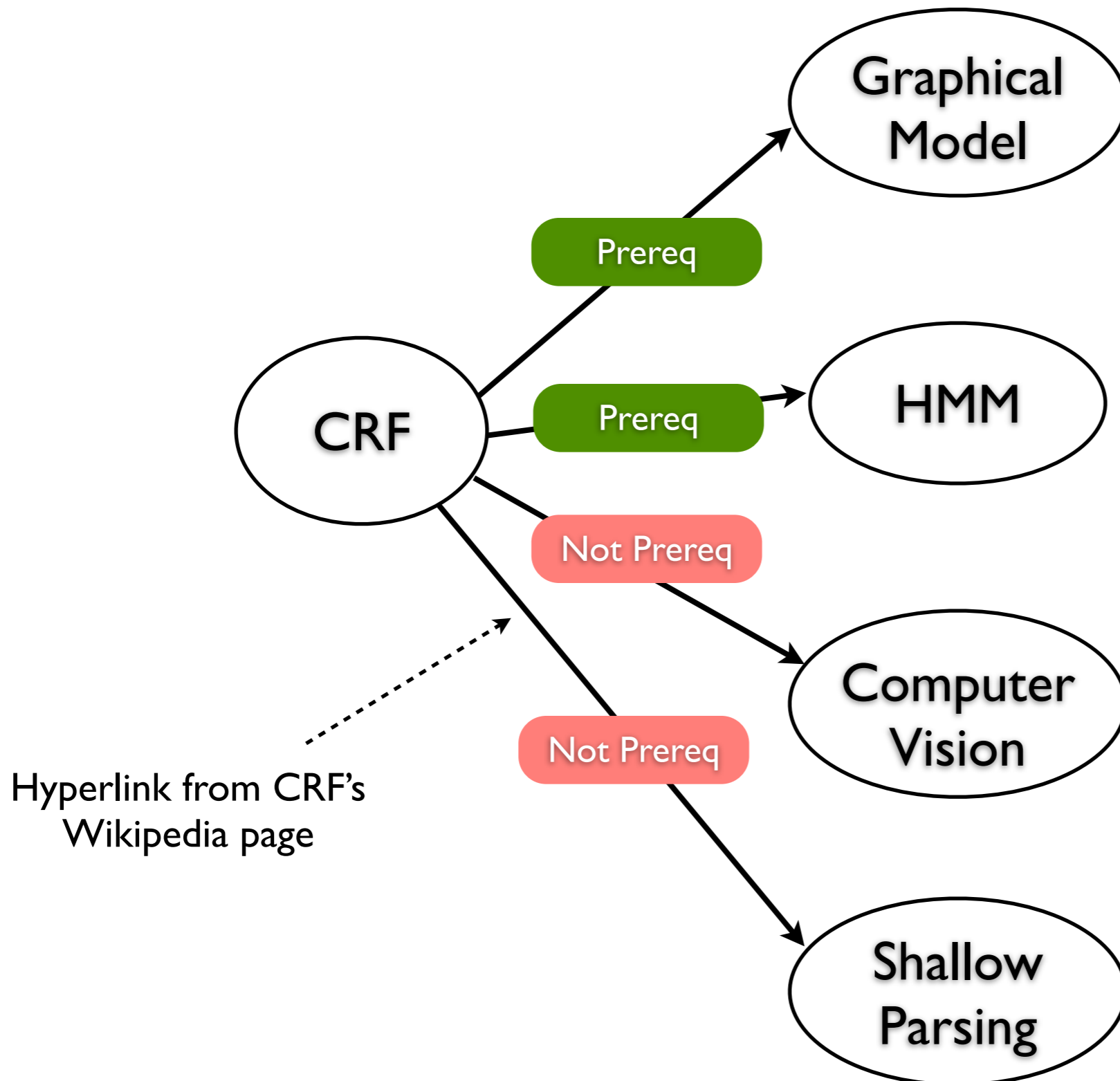
- Plan over prerequisite structure to connect user's current knowledge with target topic

Prerequisites  $\subseteq$  Hyperlinks

# Prerequisites $\subseteq$ Hyperlinks



# Prerequisites $\subseteq$ Hyperlinks



# Prerequisites $\subseteq$ Hyperlinks

Graphical Model

Prereq

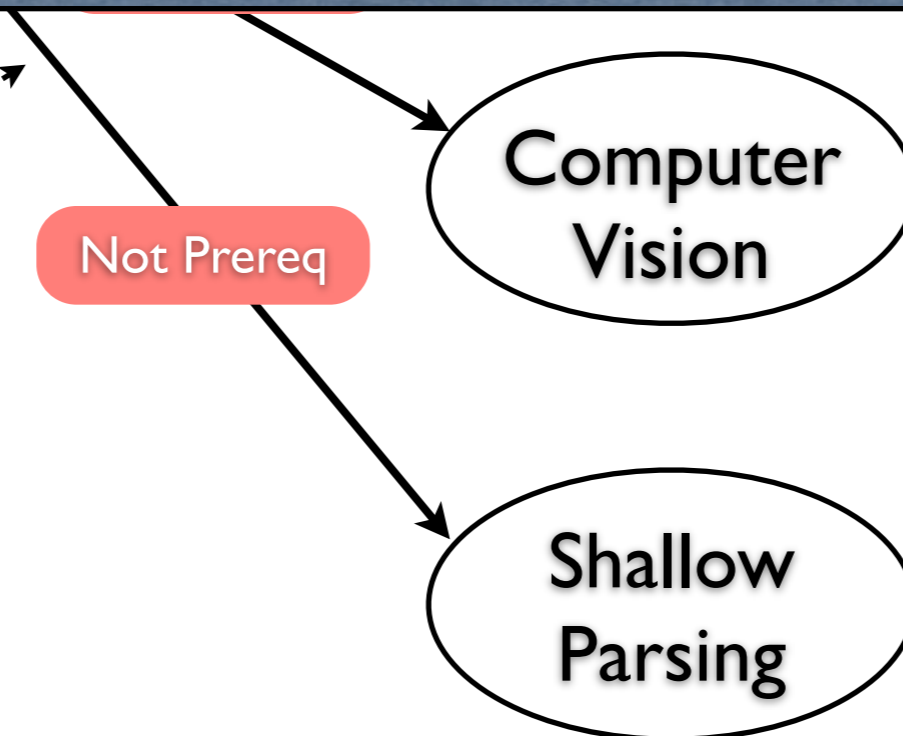
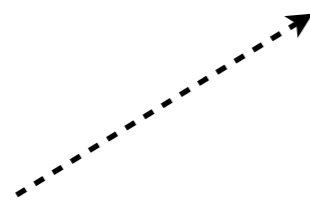
- All hyperlinks are not prerequisites.
- Given a hyperlink, classify whether it is a prerequisite

Computer Vision

Not Prereq

Shallow Parsing

Hyperlink from CRF's  
Wikipedia page





# Outline

- Motivation
- Prerequisite Classification
  - Data Preprocessing
  - Classifier
  - Prerequisite Judgements using MTurk
  - Features
- Experiments
- Conclusion

# Data Preprocessing: Target Concepts

# Data Preprocessing: Target Concepts

| Target Concept          | #Nodes | #Edges  | #Edits    |
|-------------------------|--------|---------|-----------|
| Global Warming          | 19,170 | 501,608 | 1,490,967 |
| Meiosis                 | 19,811 | 444,100 | 880,684   |
| Newton's Laws of Motion | 15,714 | 436,035 | 795,988   |
| Parallel Postulate      | 14,966 | 363,462 | 858,785   |
| Public-key cryptography | 16,695 | 371,104 | 1,003,181 |

Table 1: Target concepts used in the experiments.

Sample up to 400 edges from each subgraph

# Gold Prerequisite Judgements using Mechanical Turk

# Gold Prerequisite Judgements using Mechanical Turk

## Determine the Prerequisite Concept

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand [Acceleration](#), one should first read and understand [Velocity](#), as Velocity is a prerequisite for Acceleration. However, the reverse is not true. Similarly, [Atom](#) is a prerequisite for [molecule](#).

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.

# Gold Prerequisite Judgements using Mechanical Turk

## Determine the Prerequisite Concept

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand [Acceleration](#), one should first read and understand [Velocity](#), as Velocity is a prerequisite for Acceleration. However, the reverse is not true. Similarly, [Atom](#) is a prerequisite for [molecule](#).

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.

Select only ONE of the following:

- Is [Chromosome](#) a prerequisite for [Spirochaete](#)?
- Is [Spirochaete](#) a prerequisite for [Chromosome](#)?
- Related but one is not a prerequisite of the other.
- The two concepts are unrelated.
- I do not know the answer.

Please provide comments or feedback below:

Submit

# Gold Prerequisite Judgements using Mechanical Turk

## Determine the Prerequisite Concept

In many situations, we need to read and understand the main concept in one Wikipedia page before we can understand the main concept in another Wikipedia page. For example, in order to understand [Acceleration](#), one should first read and understand [Velocity](#), as Velocity is a prerequisite for Acceleration. However, the reverse is not true. Similarly, [Atom](#) is a prerequisite for [molecule](#).

- Below, you will be asked to compare two concepts from Wikipedia, and determine whether one is a prerequisite for the other. Clicking on a concept will take you to its corresponding Wikipedia page.
- If you are unsure about the concepts, or whether one is a prerequisite of the other, please select "I don't know the answer" option.

Select only ONE of the following:

- Is [Chromosome](#) a prerequisite for [Spirochaete](#)?
- Is [Spirochaete](#) a prerequisite for [Chromosome](#)?
- Related but one is not a prerequisite of the other.
- The two concepts are unrelated.
- I do not know the answer.

Added based on  
Turker feedback

Please provide comments or feedback below:

Submit

# Agreement among Turkers



# Agreement among Turkers

| Domain             | Time (s) / Evaluation | Worker / HIT | # HITs | $\kappa$ |
|--------------------|-----------------------|--------------|--------|----------|
| Meiosis            | 38                    | 3            | 400    | 0.50     |
| Public-key Cryp.   | 26                    | 3            | 200    | 0.63     |
| Parallel Postulate | 41                    | 3            | 200    | 0.55     |
| Newton's Laws      | 20                    | 5            | 400    | 0.47     |
| Global Warming     | 14                    | 5            | 400    | 0.56     |
| Average            | 27.8                  | -            | -      | 0.54     |

Table 2: Statistics about the Gold-standard data prepared using Amazon Mechanical Turk. Also shown are the averaged  $\kappa$  statistics-based inter-annotator agreement in each domain. The last row corresponds to the  $\kappa$  value averaged across all five domains.

# Agreement among Turkers

| Domain             | Time (s) / Evaluation | Worker / HIT | # HITs | $\kappa$ |
|--------------------|-----------------------|--------------|--------|----------|
| Meiosis            | 38                    | 3            | 400    | 0.50     |
| Public-key Cryp.   | 26                    | 3            | 200    | 0.63     |
| Parallel Postulate | 41                    | 3            | 200    | 0.55     |
| Newton's Laws      | 20                    | 5            | 400    | 0.47     |
| Global Warming     | 14                    | 5            | 400    | 0.56     |
| Average            | 27.8                  | -            | -      | 0.54     |

Table 2: Statistics about the Gold-standard data prepared using Amazon Mechanical Turk. Also shown are the averaged  $\kappa$  statistics-based inter-annotator agreement for each domain. The last row corresponds to the averaged across all five domains.

Total cost: \$278,  
completed in a week.

# Classifier: Maximum Entropy

$$p(y|x) = \frac{\exp(\mathbf{w} \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(\mathbf{w} \cdot \phi(x, y'))}, \quad y \in Y = \{-1, +1\}$$

# Classifier: Maximum Entropy

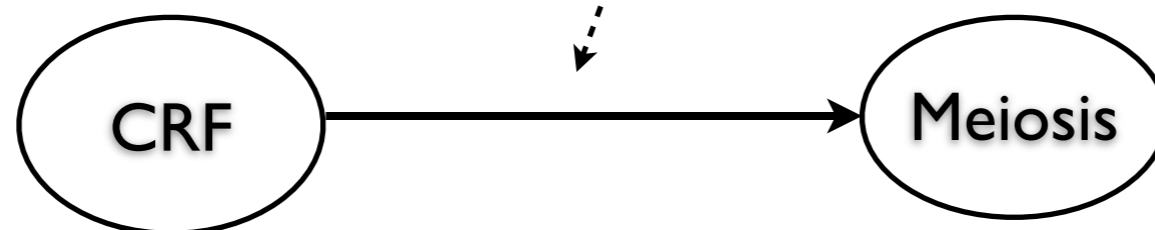
$$p(y|x) = \frac{\exp(\mathbf{w} \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(\mathbf{w} \cdot \phi(x, y'))}, \quad y \in Y = \{-1, +1\}$$

Prerequisite



# Classifier: Maximum Entropy

$$p(y|x) = \frac{\exp(\mathbf{w} \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(\mathbf{w} \cdot \phi(x, y'))}, y \in Y = \{-1, +1\}$$



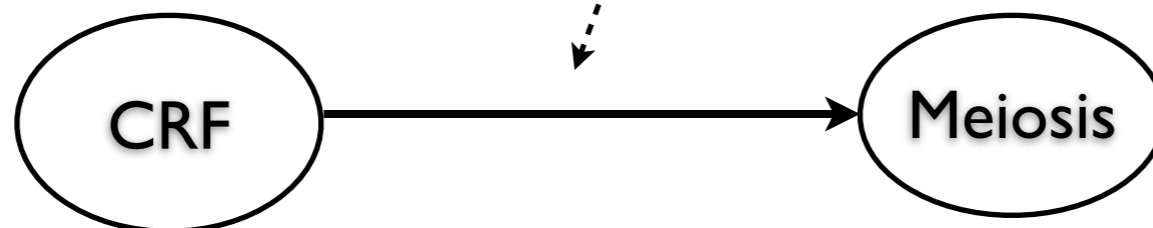
Prerequisite

# Classifier: Maximum Entropy

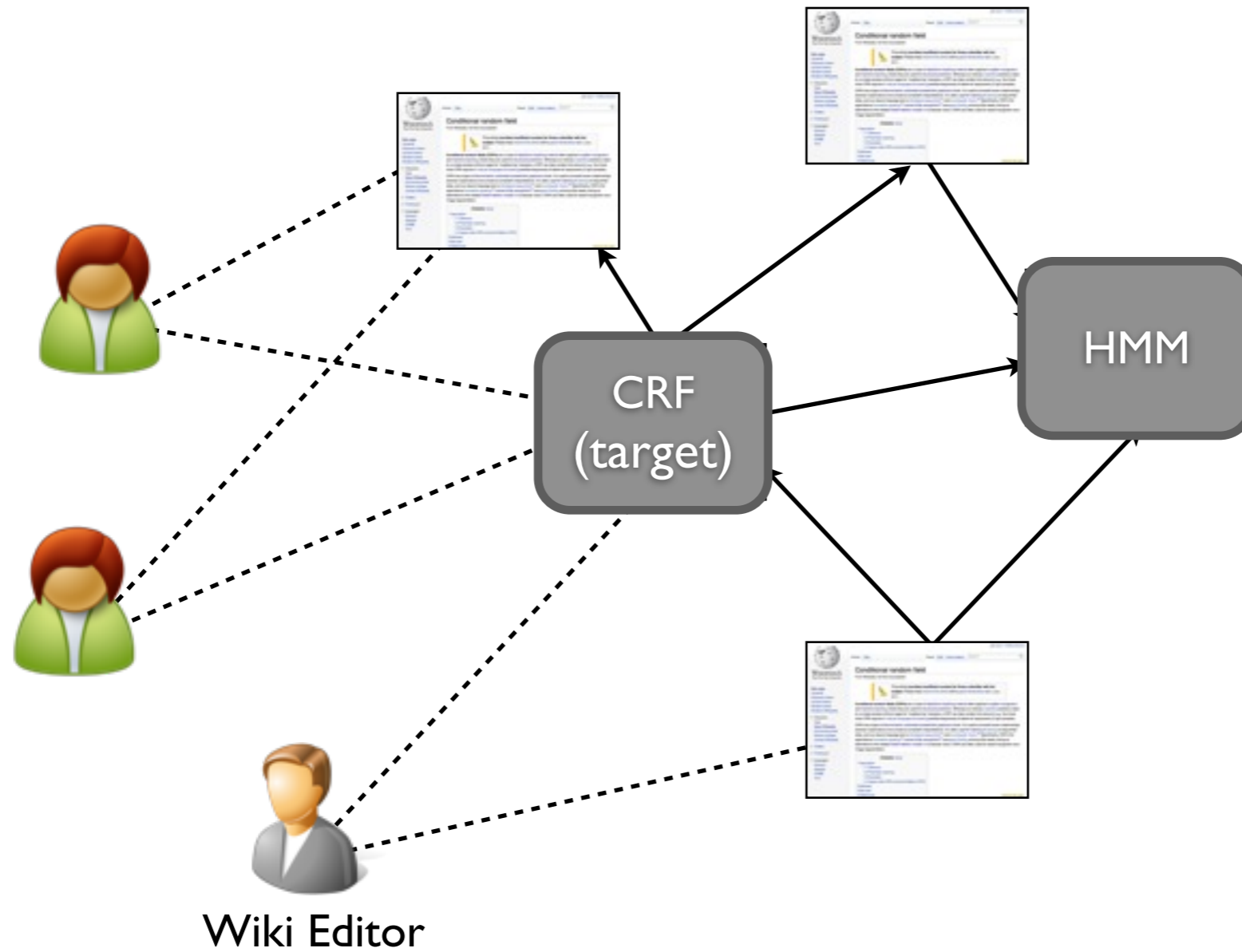
Feature Function

$$p(y|x) = \frac{\exp(\mathbf{w} \cdot \phi(x, y))}{\sum_{y' \in Y} \exp(\mathbf{w} \cdot \phi(x, y'))}, y \in Y = \{-1, +1\}$$

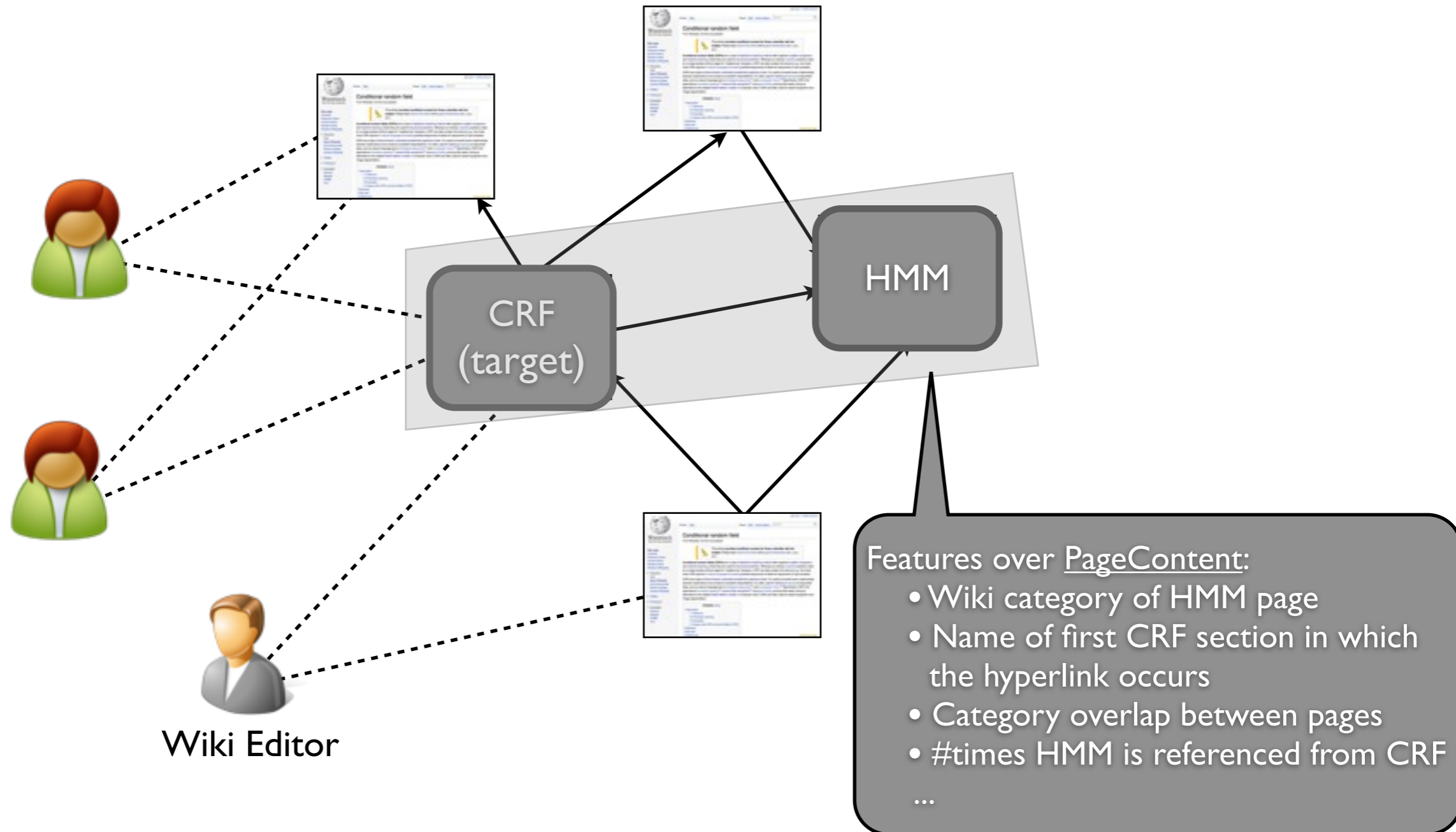
Prerequisite



# Features



# Features

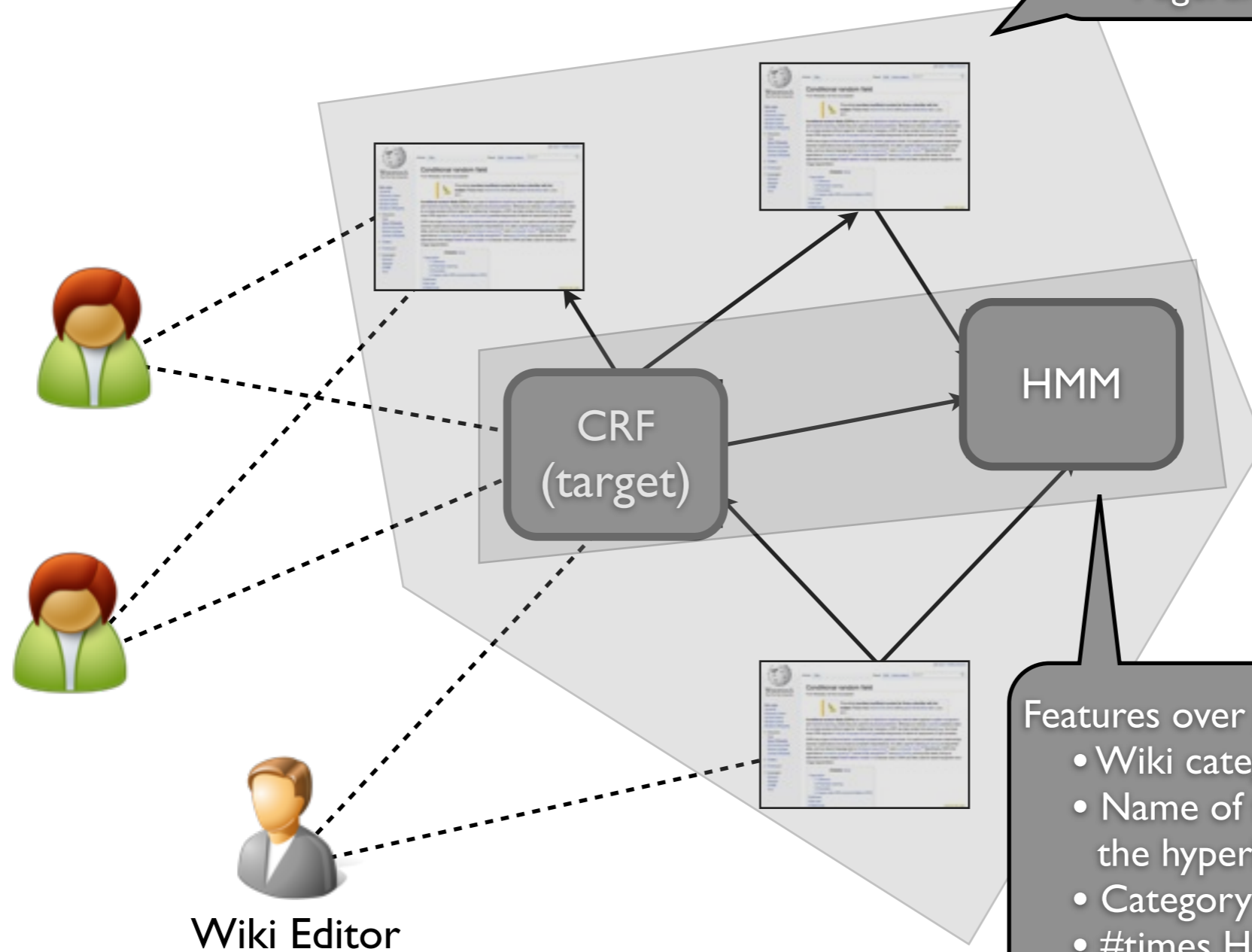




# Features

Features over Hyperlink Graph:

- Random Walk with Restart (RWR) score of a topic
- Pagerank score of source
- Pagerank score of target



Features over PageContent:

- Wiki category of HMM page
- Name of first CRF section in which the hyperlink occurs
- Category overlap between pages
- #times HMM is referenced from CRF

...

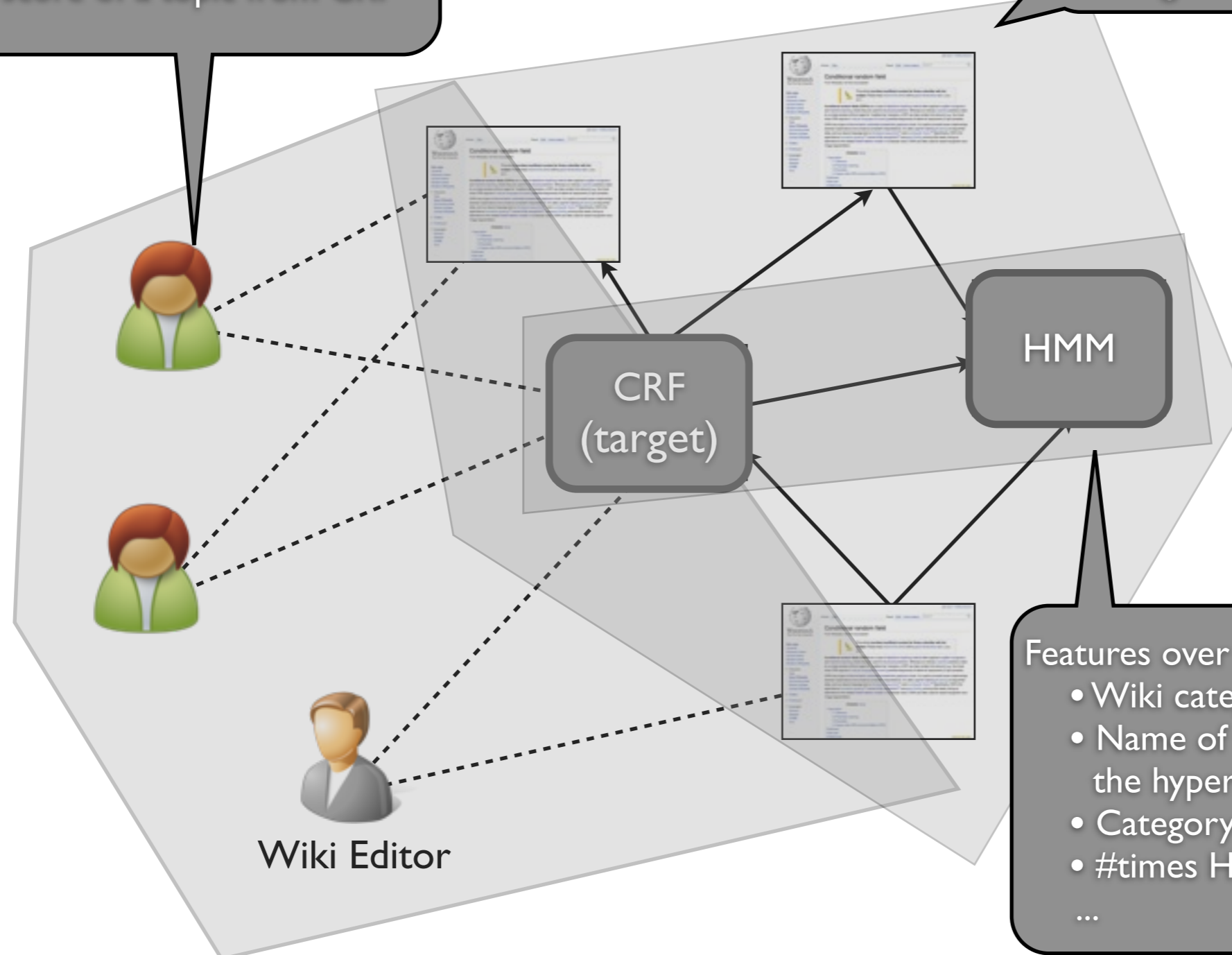
# Features

Feature over WikiEdits Graph:

- RWR score of a topic from CRF

Features over Hyperlink Graph:

- Random Walk with Restart (RWR) score of a topic
- Pagerank score of source
- Pagerank score of target



Features over PageContent:

- Wiki category of HMM page
- Name of first CRF section in which the hyperlink occurs
- Category overlap between pages
- #times HMM is referenced from CRF

...

# Experiments

# Experiments

- Can we train a classifier for prerequisite classification?

# Experiments

- Can we train a classifier for prerequisite classification?
- How effective is out-of-domain training?

# Experiments

- Can we train a classifier for prerequisite classification?
- How effective is out-of-domain training?
- What are the effects of different features?

# Experiments

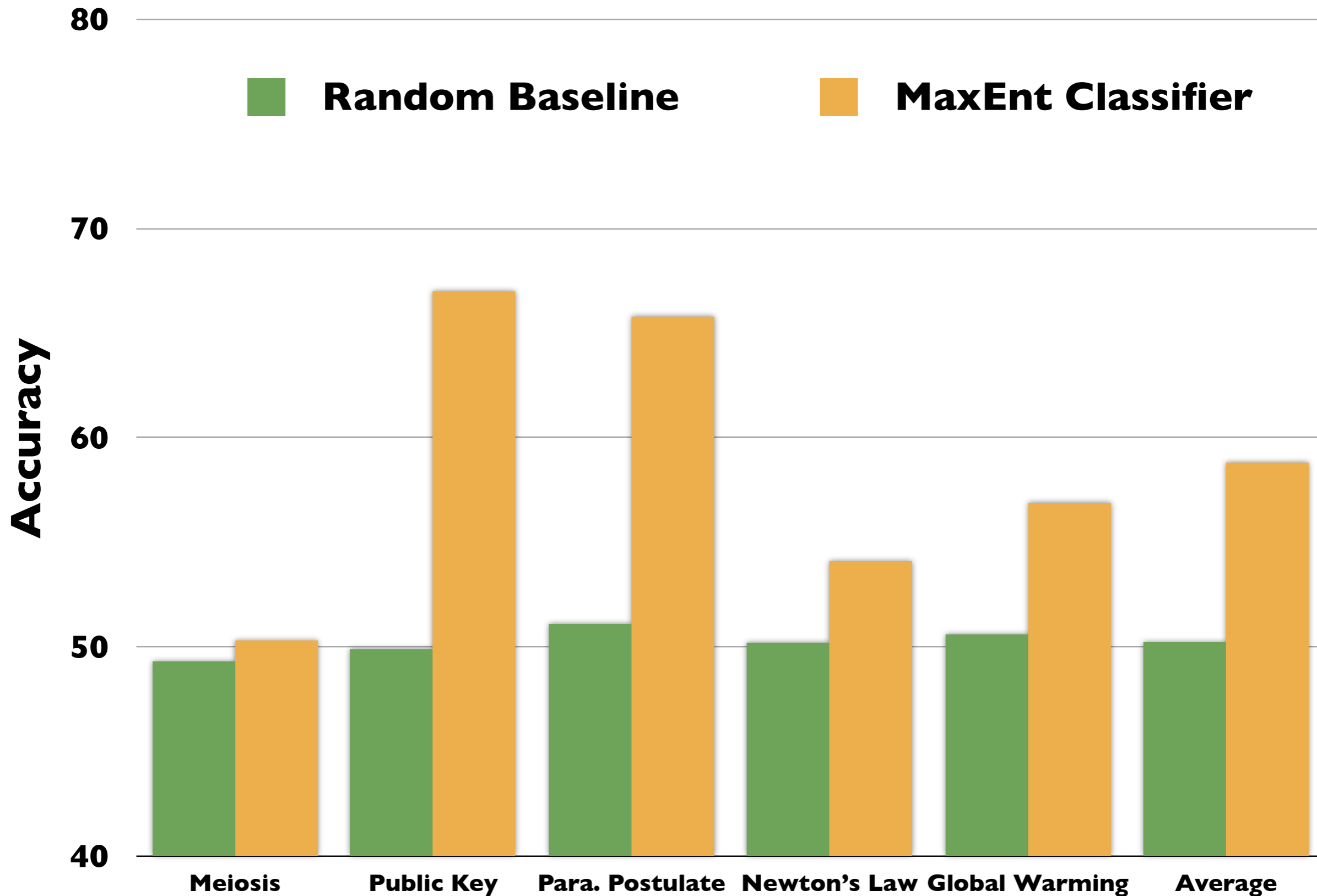
- Can we train a classifier for prerequisite classification?
- How effective is out-of-domain training?
- What are the effects of different features?

All evaluations in leave-one-target-out setting

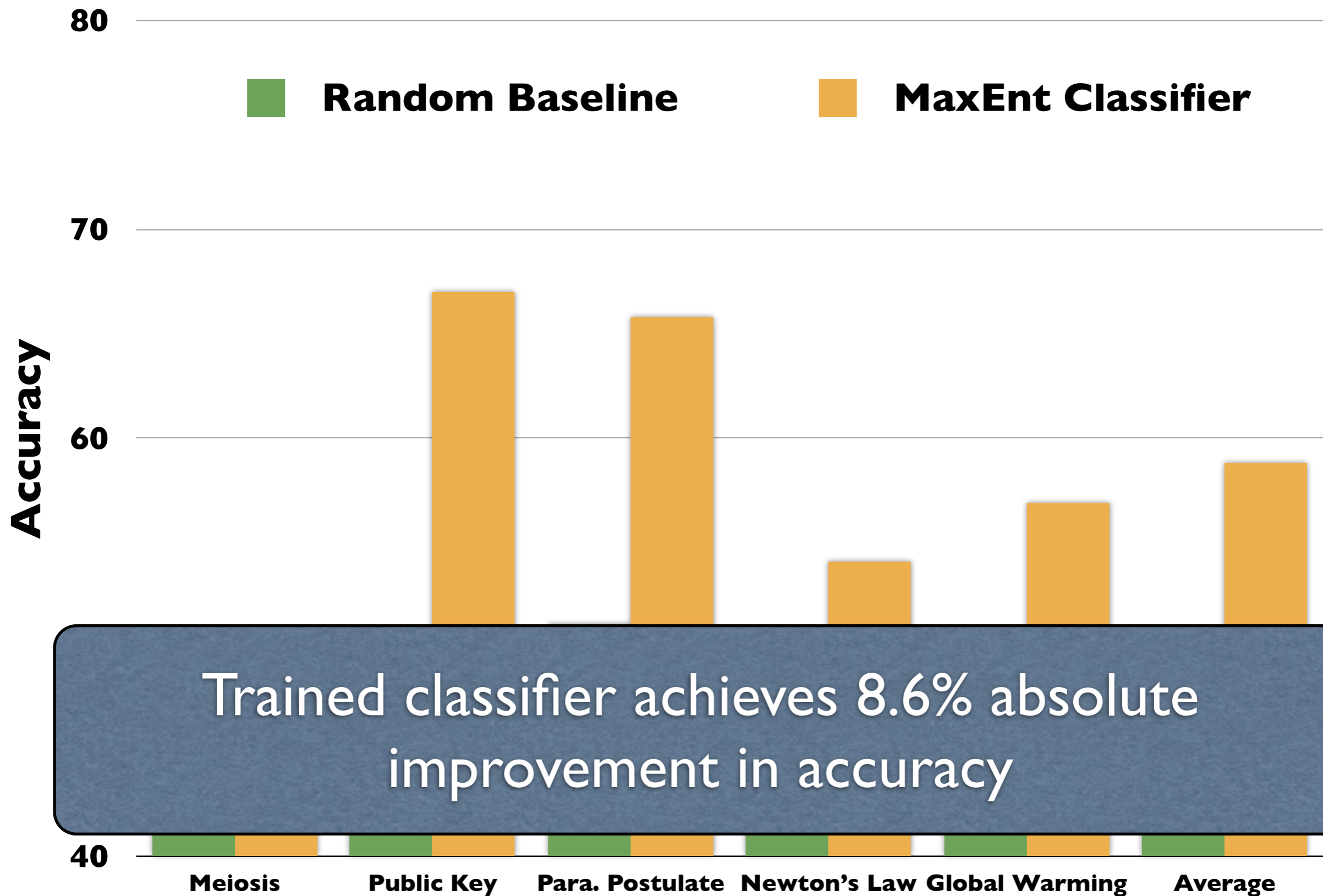




# Performance Comparison for Prerequisite Prediction



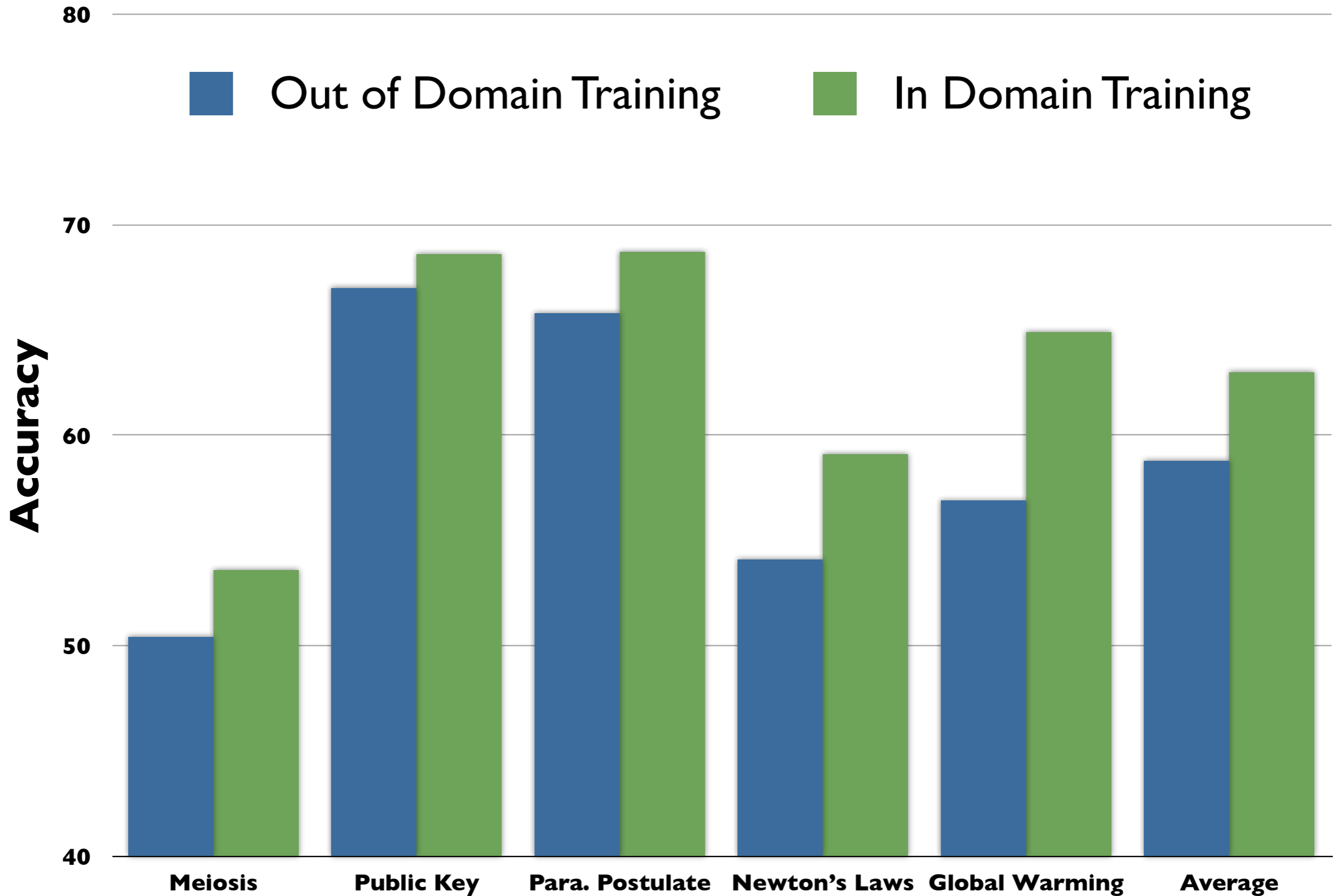
# Performance Comparison for Prerequisite Prediction



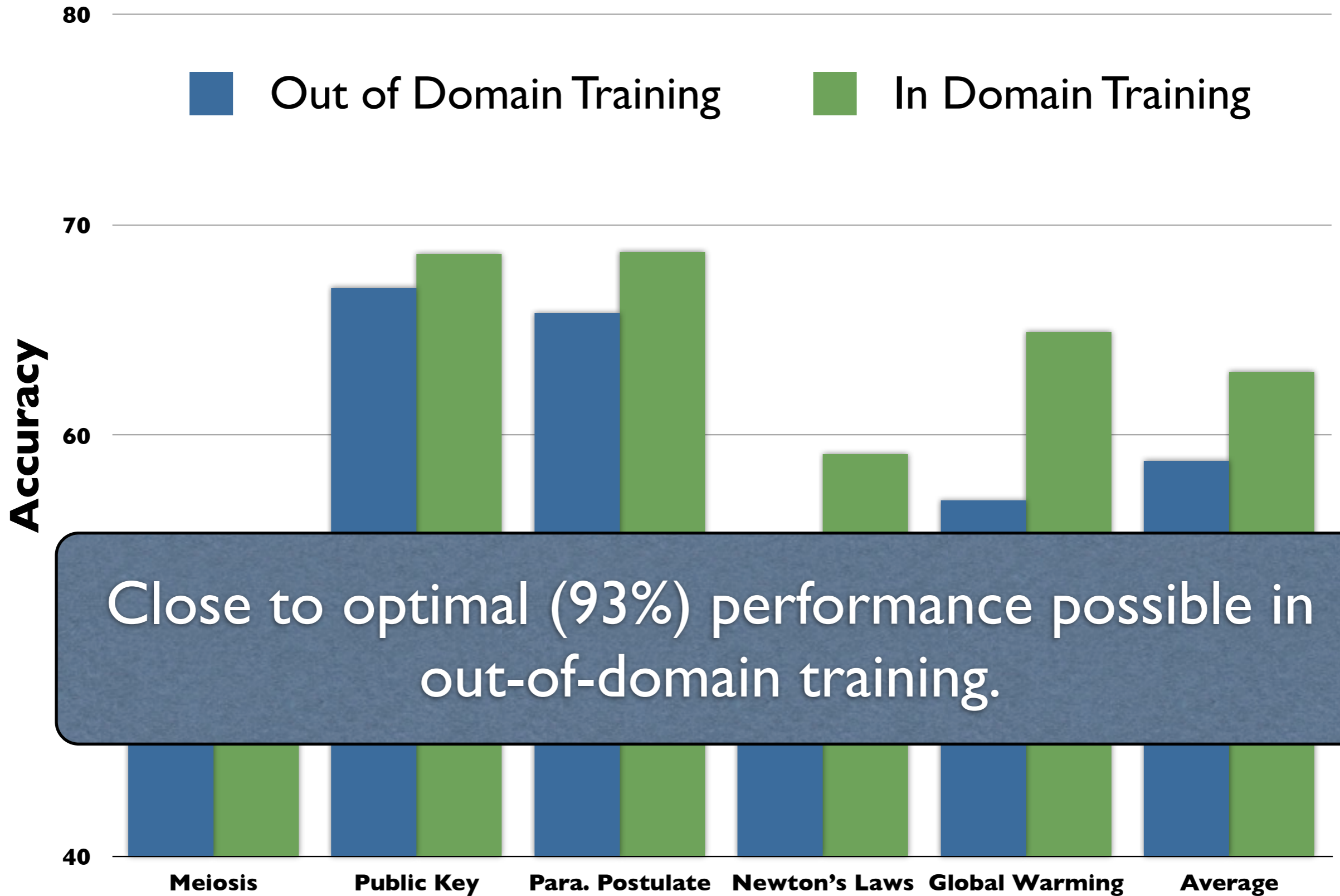
Trained classifier achieves 8.6% absolute improvement in accuracy



# Effect of Out of Domain vs In Domain Training



# Effect of Out of Domain vs In Domain Training



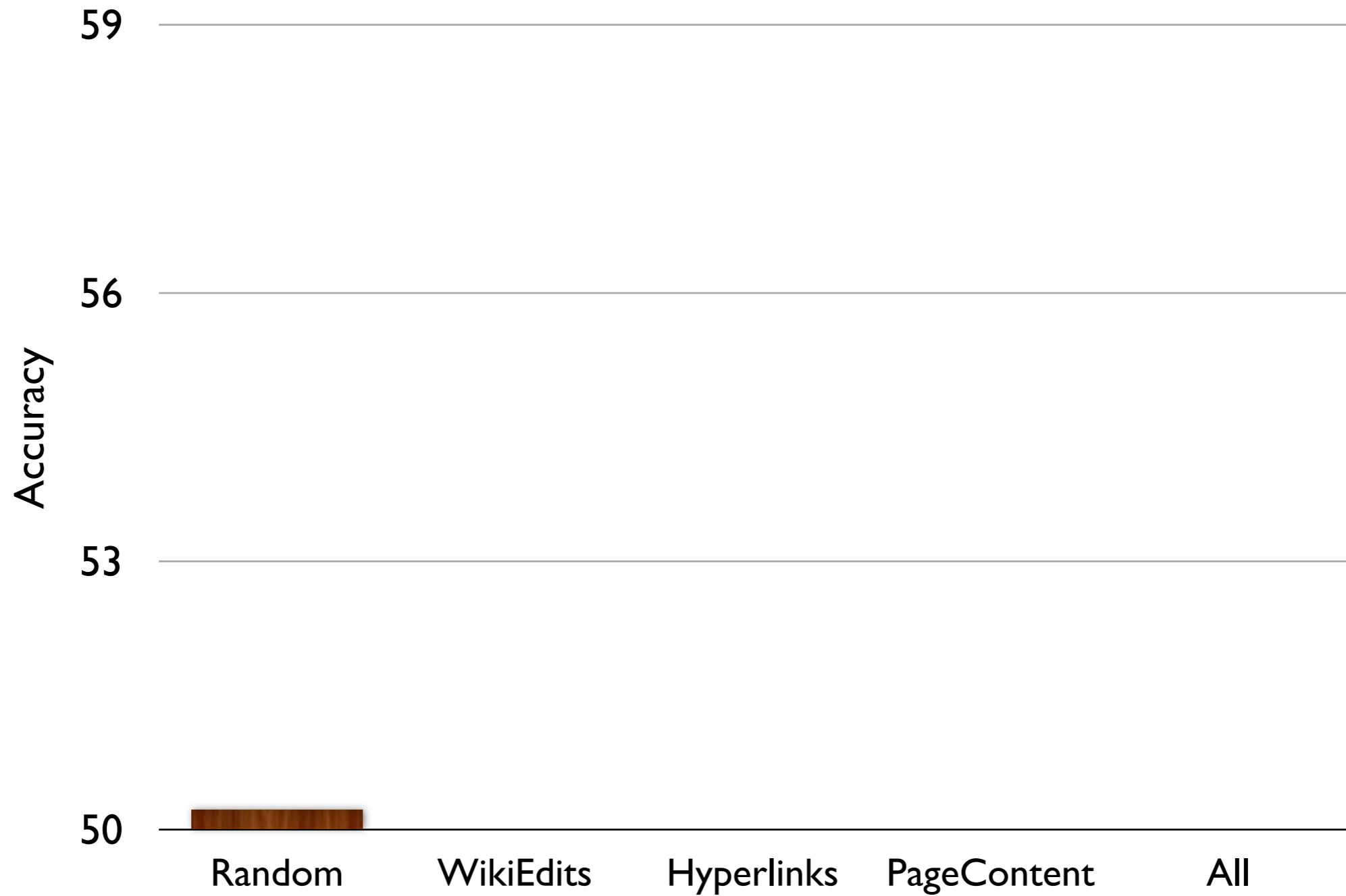
Close to optimal (93%) performance possible in out-of-domain training.

# Effect of Different Features

# Effect of Different Features

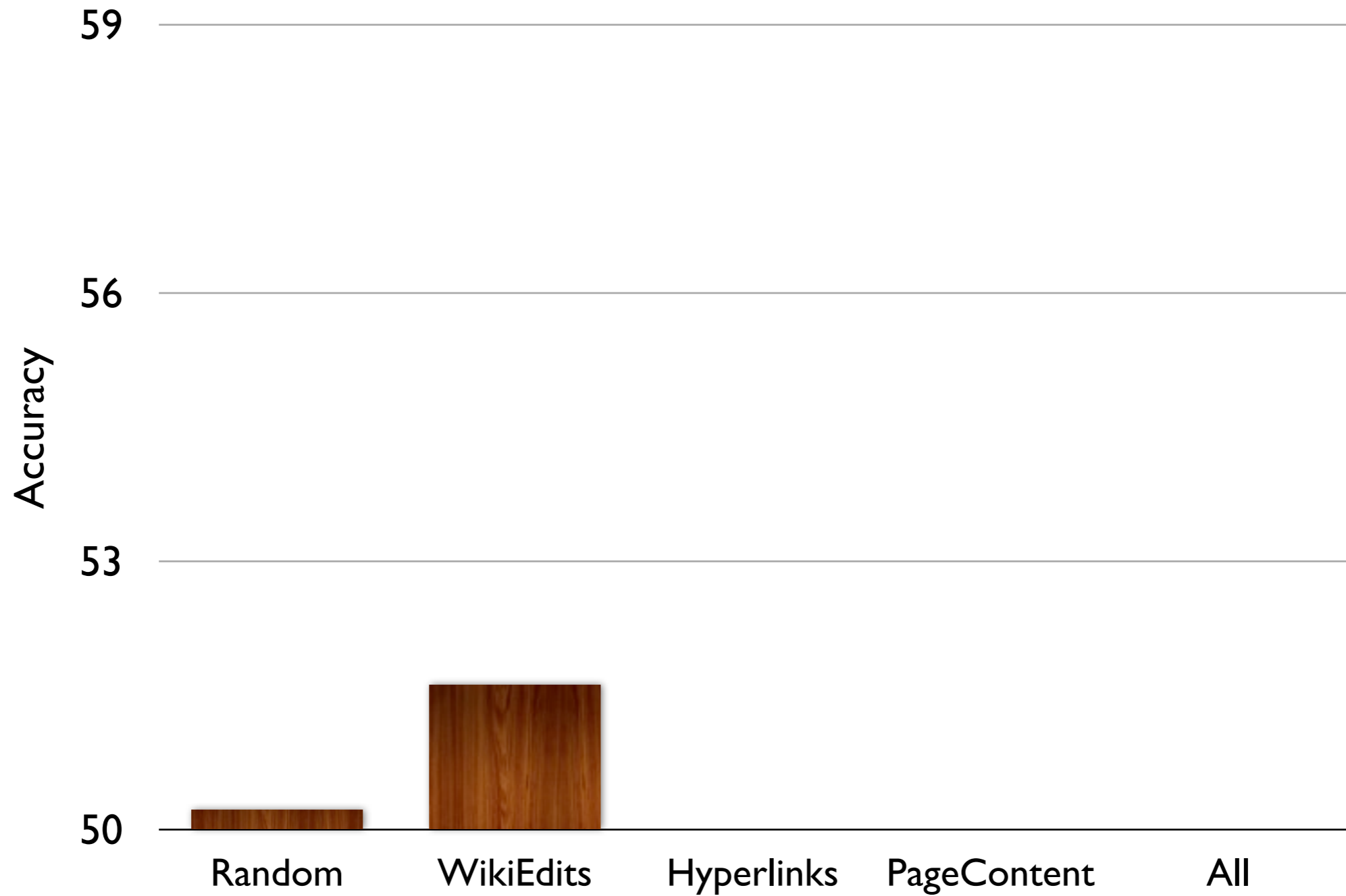


# Effect of Different Features

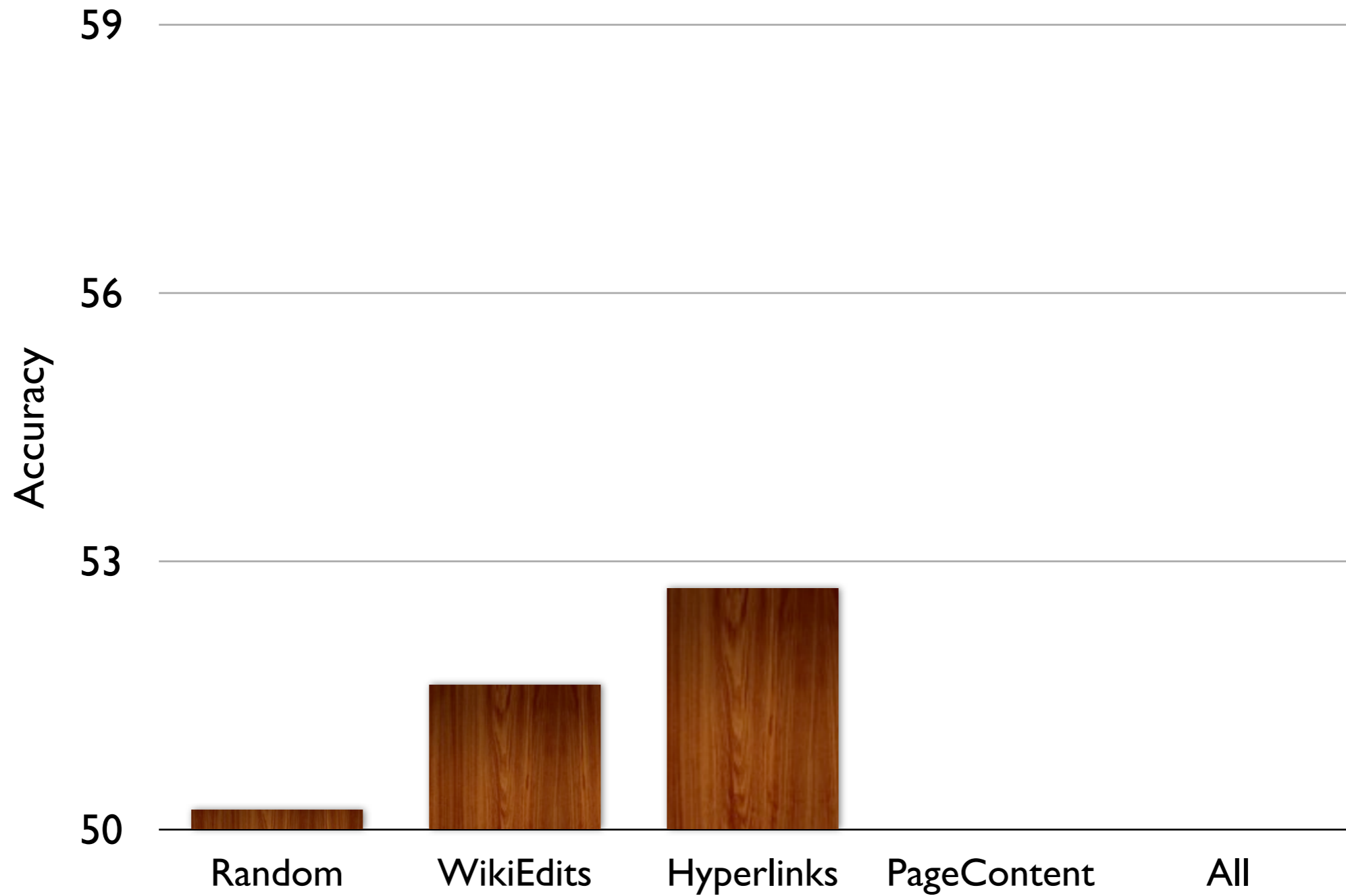




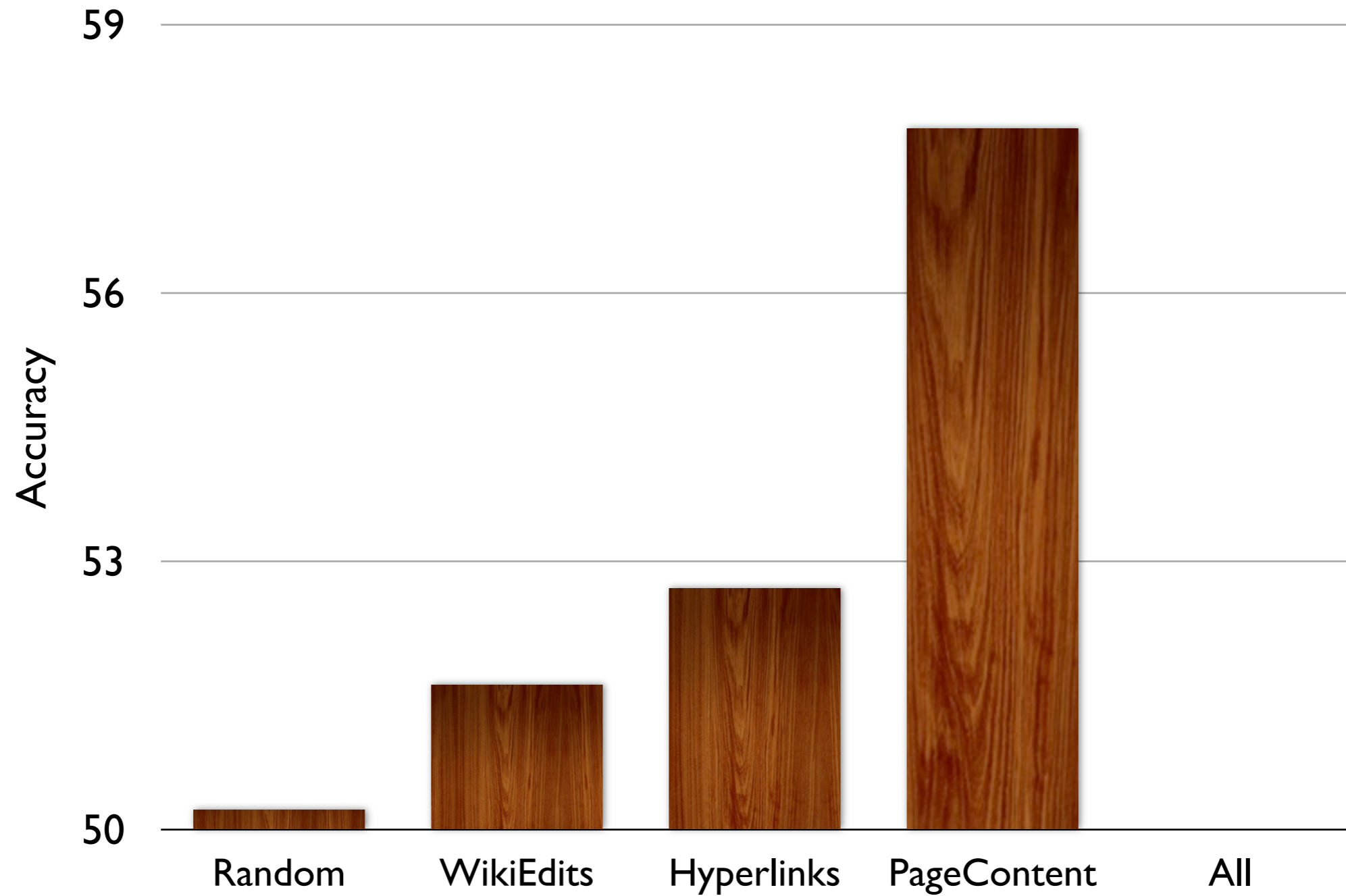
# Effect of Different Features



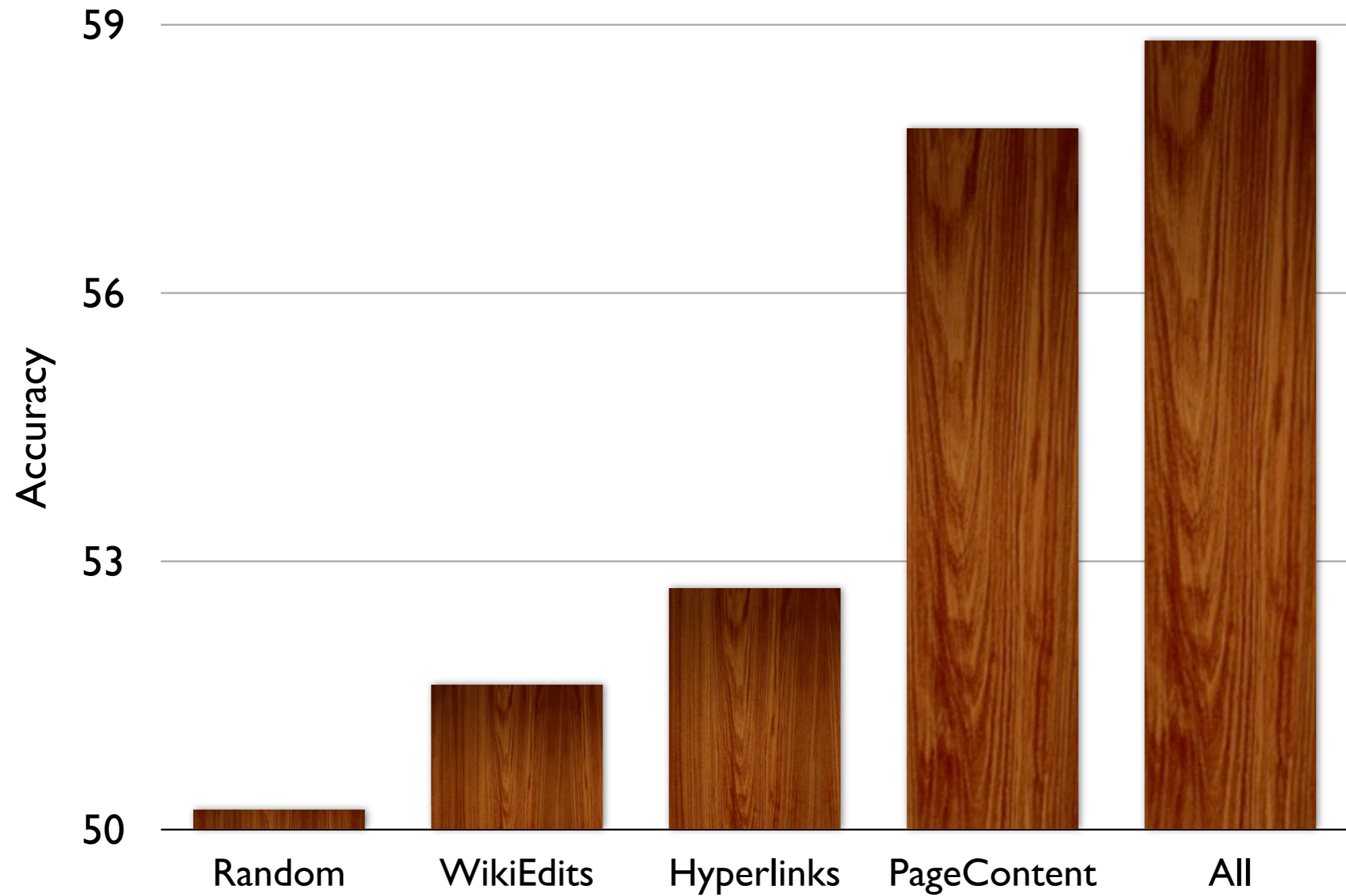
# Effect of Different Features



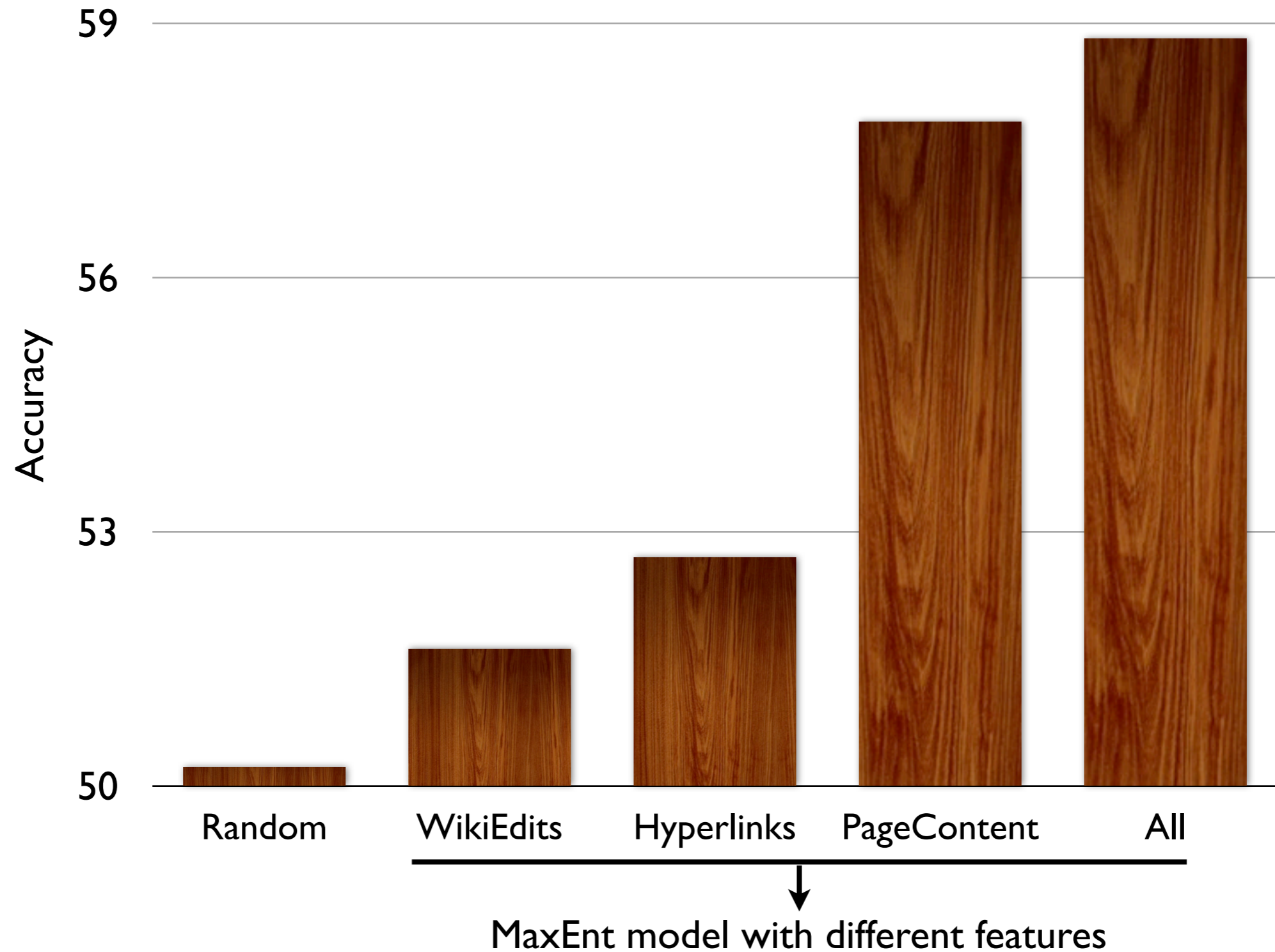
# Effect of Different Features



# Effect of Different Features



# Effect of Different Features



# Incorrect Classifications

# Incorrect Classifications

- Mass is not a prerequisite for Physics
- Quantum mechanics is not a prerequisite for Bohr-Einstein debates
- Global warming is a prerequisite for Nitrous Oxide
- Carbon Dioxide is not a prerequisite for Carbon Sequestration

# Examples of Discriminative Features



# Examples of Discriminative Features

- Is Target title a substring of Source title?

# Examples of Discriminative Features

- Is Target title a substring of Source title?
- Is there category overlap between source and target pages?

# Examples of Discriminative Features

- Is Target title a substring of Source title?
- Is there category overlap between source and target pages?
- Identity of category overlap between pages.

# Examples of Discriminative Features

- Is Target title a substring of Source title?
- Is there category overlap between source and target pages?
- Identity of category overlap between pages.
- Identity of source section from where target is linked.

# Conclusion

# Conclusion

- Novel task: prerequisite structure prediction
  - Demonstrated that relatively reliable features exist

# Conclusion

- Novel task: prerequisite structure prediction
  - Demonstrated that relatively reliable features exist
- Future work
  - feature engineering
  - controlled data collection

# Conclusion

- Novel task: prerequisite structure prediction
  - Demonstrated that relatively reliable features exist
- Future work
  - feature engineering
  - controlled data collection
- Ongoing work
  - full comprehension plan generation using predicted prerequisite structure



**Thank You!**