UW-COSMOS

Shanan Peters, Miron Livny, Shivaram Venkataraman
Theodoros Rekatsinas

Table Retrieval and Co-References
Retrieve artifacts from publications applicable to model and link as meta-data

Table 1: Parameters used in Discrete Stochastic Compartmental Model

Parameter	Guilford County	North Carolina
Import Date	2021-01-28	2021-01-23
Reproduction Number	0.97(0.88-1.03)	0.96(0.89-1.02)
Population	545,348	10,630,691
Susceptible	454,109	8,626,950
Exposed	1,964	31,944
Infected	3,928	63,888
Recovered (Natural Immunity)	76,490	1,669,565
Vaccinated	8,857	238,344
Vaccination Rate (Doses per Day)	825	21,082
Vaccine Efficiency	95%	95%
Vaccine Uptake	100%	100%
Variant Transmissibility Increase	0%, 50%, 80%	0%, 50%, 80%

Rapid Impact Analysis of B 1.1.7 Variant on the Spread of SARS-CoV-2 in North Carolina
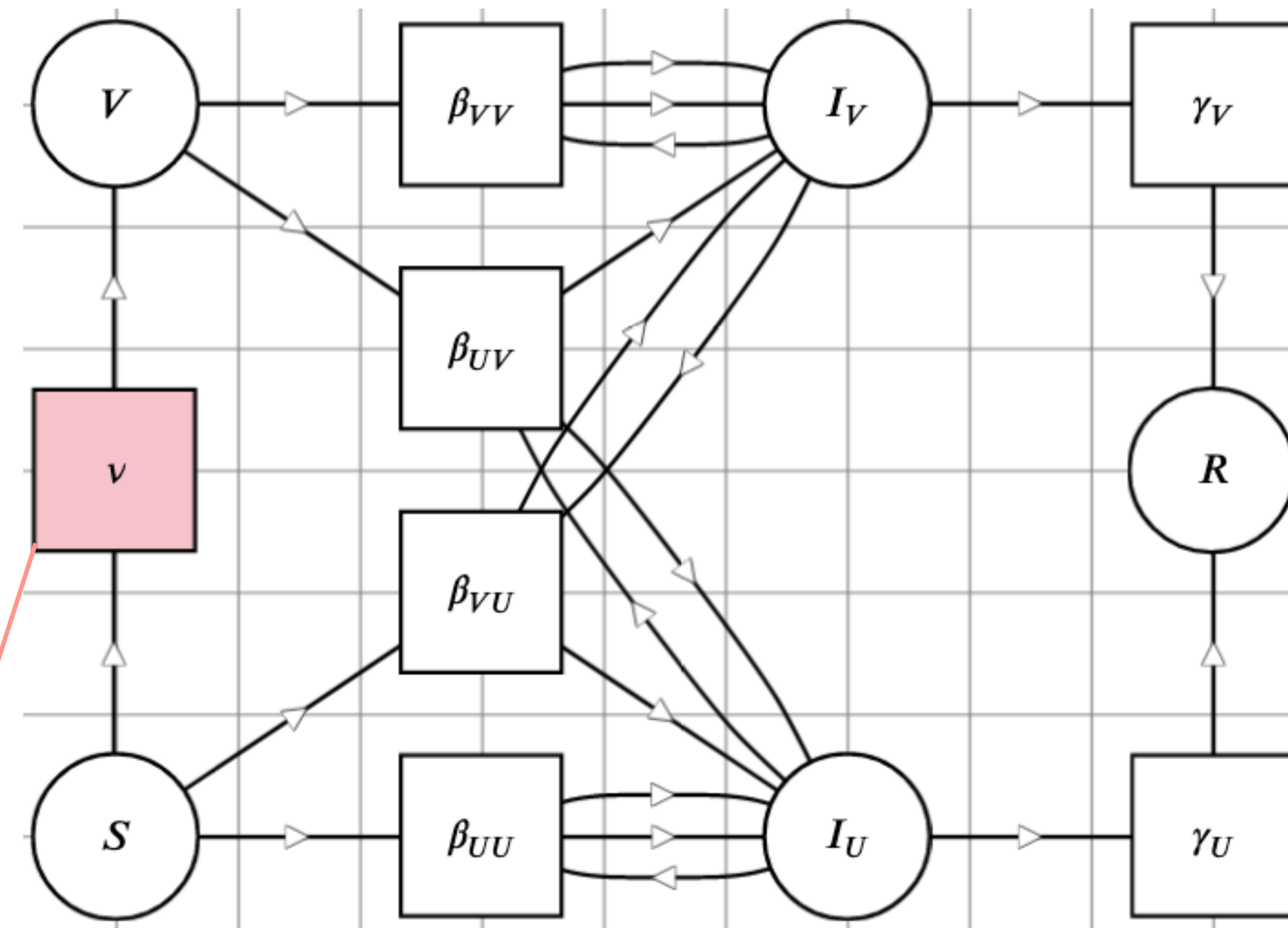
Michael DeWitt

medRχiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Context search on node returns tables from literature with vaccination data.

Intervention Parameter: daily vaccination rate

SVIvR model representing vaccine intervention

UW-COSMOS

Table Retrieval and Co-References
Incorporate Body-Text Content into Table and Figure Retrieval

Table 1: Parameters used in Discrete Stochastic Compartmental Model

Parameter		
Import Date		
Reproduction		
Population		
Susceptible		
Exposed		
Infected		
Recovered (N		
Vaccinated		
Vaccination		
Vaccine Effic		
Vaccine Uptake	100%	100%
Variant Transmissibility Increase	0%, 50%, 80%	0%, 50%, 80%

Nine scenarios were evaluated for North Carolina and Guilford County. These scenarios included the 10%, 50%, and 90% quantile estimates for the effective reproduction number on the incubation time adjusted import date of B1.1.7 estimated using the EpiNow2 package.[7] The increase in transmissibility was modeled as 0% reflecting no increase, 50% increase, and 80% increase.[4,5,14] Based on the latest findings from Davies, the increase in transmissibility could be as high as 82% with the 95% credible interval including 106%.[6]. The scenarios are shown in Table 1.

Rapid Impact Analysis of B 1.1.7 Variant on the Spread of SARS-CoV-2 in North Carolina

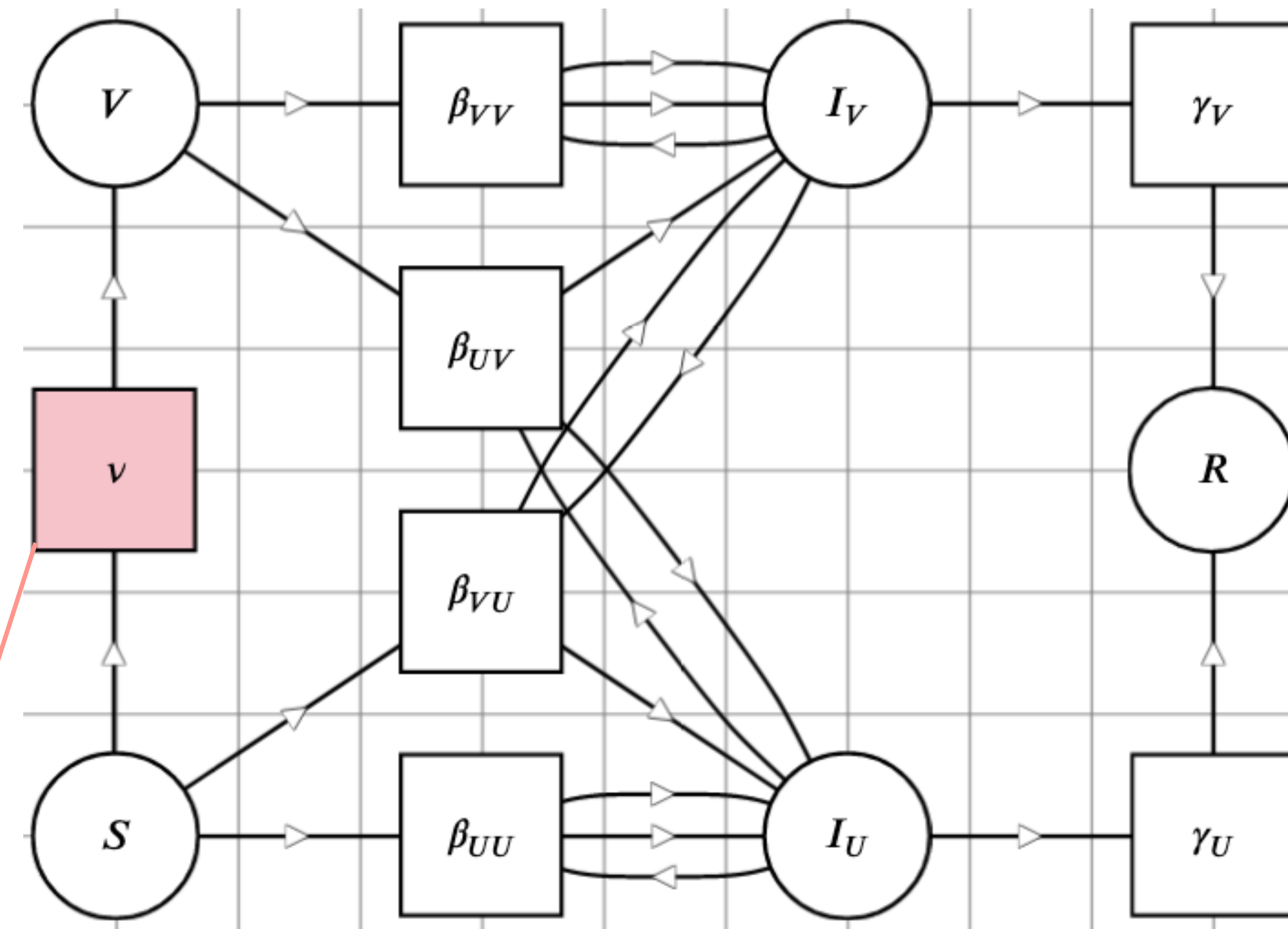Michael DeWitt

doi: https://doi.org/10.1101/2021.02.07.21251291

medRxiv
THE PREPRINT SERVER FOR HEALTH SCIENCES

Retrieval now returns additional context from full-text, providing richer domain explanations

SVIvR model representing vaccine intervention

Intervention Parameter: daily vaccination rate

UW-COSMOS

Table Retrieval and Co-References
Automated table reading turns PDF into useable data

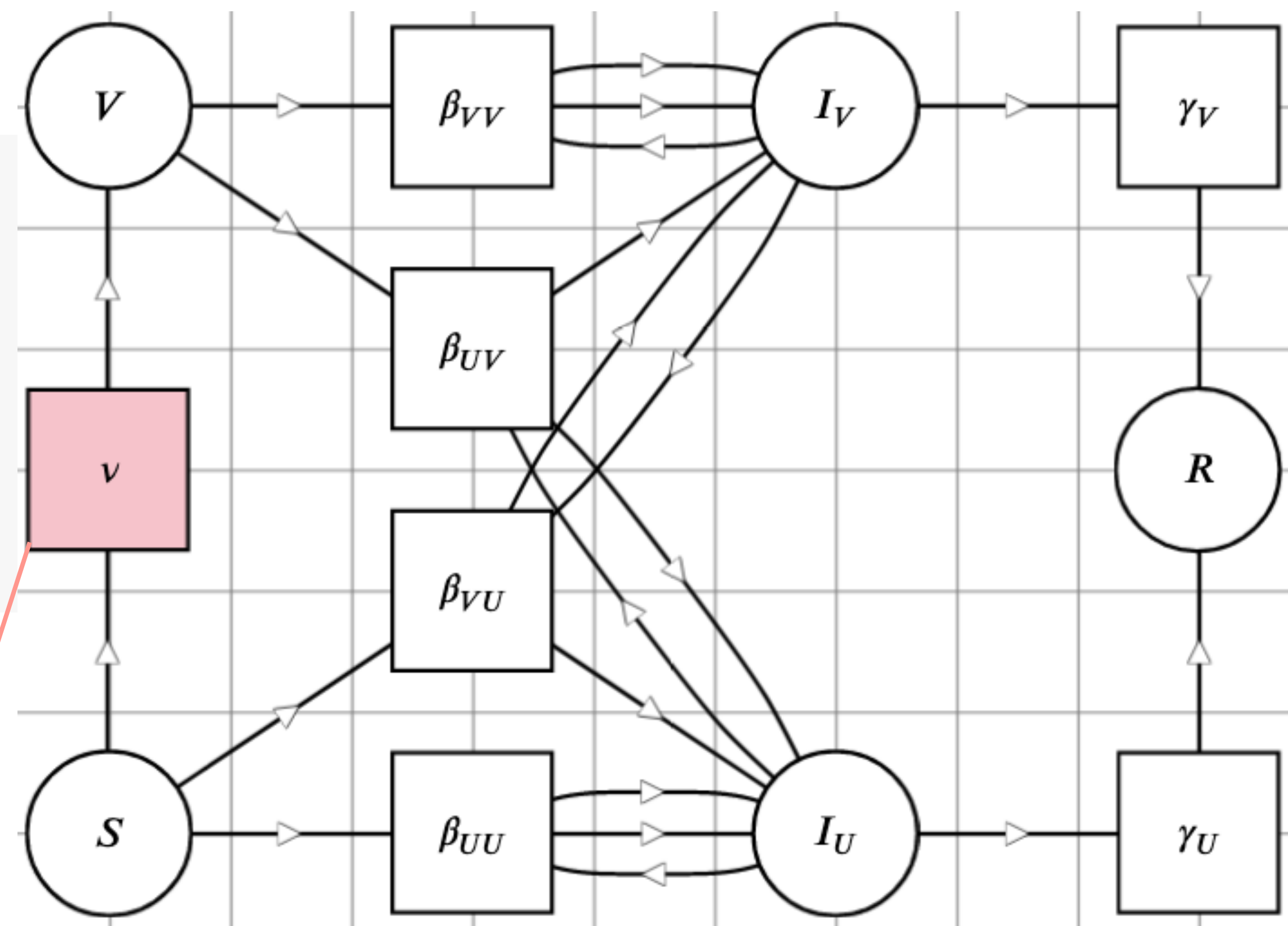**SVIvR model representing vaccine intervention**

JSON-representation ready for ingestion

```
    "cls" : "Table",
    "dataset_id" : "documents_5Feb",
    "content" : "Parameter Guilford County North Carolina Import Date 2021-01-28 2021-01-23
Reproduction Number 0.97(0.88-1.03) 0.96(0.8
9-1.02) Population 545,348 10,630,691 Susceptible 454,109 8,626,950 Exposed 1,964 31,944
Infected 3,928 63,888 Recovered (Natural Immuni
ty) 76,490 1,669,565 Vaccinated 8,857 238,344 Vaccination Rate (Doses per Day) 825 21,082
Vaccine Efficiency 95% 95% Vaccine Uptake 100%
 100% Variant Transmissibility Increase 0%, 50%, 80% 0%, 50%, 80%",
    "header_content" : "Table 1: Parameters used in Discrete Stochastic Compartmental Model",
    "context_from_text" : "in transmissibility could be as high as 82% with the 95% credible
interval including 106%.[6] The scenarios are shown in table 1 1Where \"removed\" could be
through recovery, death, or vaccination. The scenarios are shown in Table 1. Statistical
analysis In order scenarios are shown in Table 1. 1Where \"removed\" could be through
```

Table extraction provides direct programmatic access to data; embeddings trained over this content.

Intervention Parameter: daily vaccination rate

Rapid Impact Analysis of B 1.1.7 Variant on the Spread of SARS-CoV-2 in North Carolina

Michael DeWitt

doi: https://doi.org/10.1101/2021.02.07.21251291

medRχiv

THE PREPRINT SERVER FOR HEALTH SCIENCES

UW-COSMOS
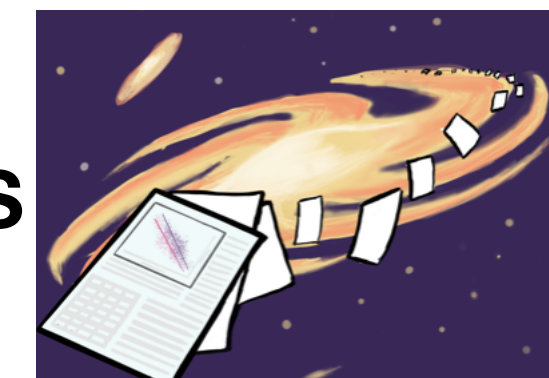
xDD and UW-COSMOS data infrastructure overview

Text-based **search and retrieval** over full text
of all 14M full texts in xDD corpus
xdd.wisc.edu/api

- api/snippets: string-based search over document full-text to retrieve document metadata and contextual snippets highlighting search matches.
- api/articles: similar to snippets, but article metadata focused
- api/journals: search for specific journals to assess coverage
- api/: additional routes and documentation

xDD API

xDD sets Subsets of xDD corpus for specific projects: xdd.wisc.edu/sets

continuous updating

xDD automated full-pub fetching 10³s per day

xDD corpus 14M full texts + PMC abstracts

periodic updating

154,777 full-text docs

sets/xdd-covid-19   **xDD-defined COVID19 doc set (CORD19, EMMAA, search-term defined)**: full content only

restrict API results to set: &dataset=xdd-covid-19

open access full-pub   abstract-only text

Europe PMC

Table, figure, equation, **retrieval** over sets
via COSMOS AI assistant (cosmos.wisc.edu)
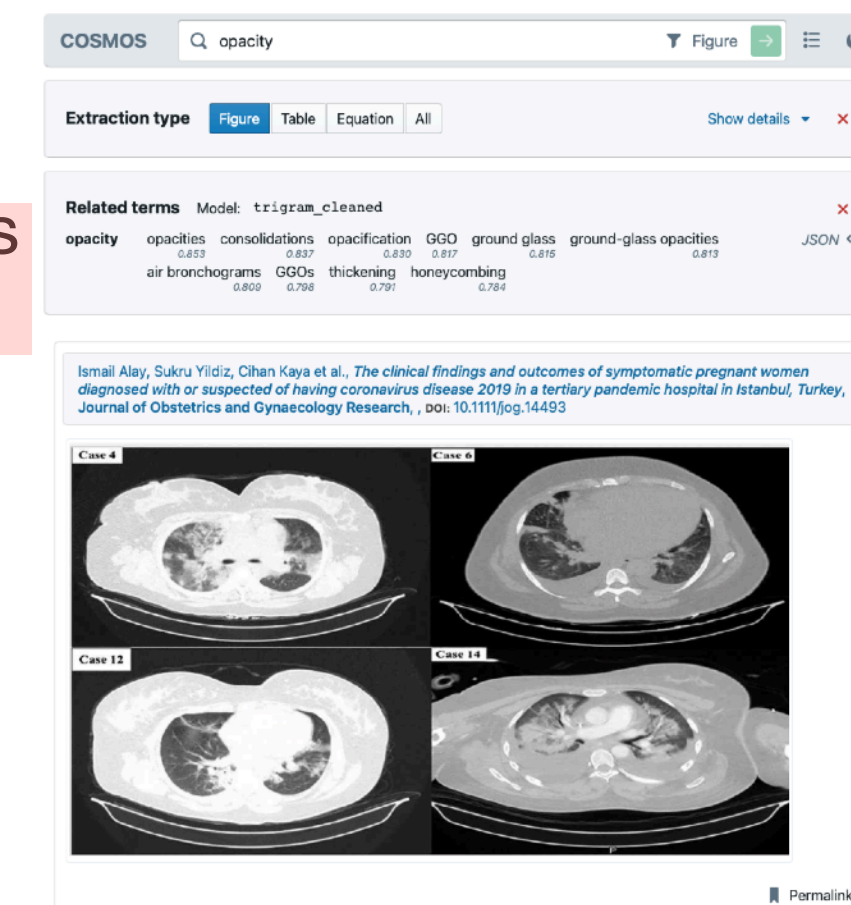**xdd.wisc.edu/sets/xdd-covid-19/cosmos/api/**

ASKE-ID   **UW COSMOS API** 🔑

- search: string-based search over document figures, tables, equations based on associated text.
- other routes available, most *require API key*

- service for generating/ retrieving metadata and links to artifacts

word2vec API

doc2vec API

xdd.wisc.edu/sets/xdd-covid-19/word2vec/api/

xdd.wisc.edu/sets/xdd-covid-19/doc2vec/api/

- trained embeddings
- several vector operations

- document-level embeddings, doc discovery based on similarity

A System for Large Scale Graph Embeddings

Anze Xie*, Anders Carlsson*, Jason Mohoney*, Roger Waleffe*

Shanan Peters#, Theodoros Rekatsinas*, Shivaram Venkataraman*

*Computer Sciences, #Geoscience

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

Our solution: Marius

- Example: human-constructed Paleobiology Database built by reading manually ~1,200 xDD scientific publications: 259,335 nodes and 3,149,181 edges

- Allows large-scale graph embedding on a single machine

- Easy-to-use config-based development framework

- An open-source system introduced in OSDI 2021

Task: Knowledge Discovery on Paleobiology Database

marius

Node label: [_____] Node property: [_____] Property value: [_____] Type of search: [Equals ▾] Results limit: [50] [Search]

Traverse by edge: [_____] Nb of layers [3] ☐ Freeze exploration ☐ Show labels [Clear]

[Get graph info] ☐ Show/hide graph info

Item Info

Key	Value

Key	Value	Property

Embeddings Inference

Edge type: [_____]

[Infer new edges]

Server Address: [localhost] Server port: [8182] Protocol: [REST ▾] Gremlin version: [3.4.* ▾] [Edit Graph]

Graph Explorer V 0.8.0

A System for Large Scale Graph Embeddings

Anze Xie*, Anders Carlsson*, Jason Mohoney*, Roger Waleffe*

Shanan Peters#, Theodoros Rekatsinas*, Shivaram Venkataraman*

*Computer Sciences, #Geoscience