

EPIDEMIOLOGÍA COMPUTACIONAL

ACTUALMENTE ME ENCUENTRO EN LA UNIVERSIDAD DE IOWA, ESTADOS UNIDOS, DOCTORÁNDOME EN EL TEMA DE EPIDEMIOLOGÍA COMPUTACIONAL, QUE CONSISTE EN LA APLICACIÓN DE MÉTODOS COMPUTACIONALES PARA OBJETIVOS DE EPIDEMIOLOGÍA Y SALUD PÚBLICA. ESTA ÁREA OTORGA UN RENOVADO VIGOR A TODOS LOS TEMAS CLÁSICOS DE LA CIENCIA DE LA COMPUTACIÓN, PASANDO POR REDES, ALGORITMOS, INTELIGENCIA ARTIFICIAL, RECUPERACIÓN DE LA INFORMACIÓN, ETC., HASTA POR INTERACCIÓN HUMANO-MÁQUINA. EN PARTICULAR INVESTIGO CÓMO OCURREN LAS EPIDEMIAS EN POBLACIONES Y EN HOSPITALES, Y CÓMO SE PUEDEN CONTROLAR ESTAS EPIDEMIAS; O EN OTRAS PALABRAS, A QUIÉN MONITOREAR, A QUIÉN VACUNAR, CUÁNDO VACUNAR, CÓMO IDENTIFICAR A LOS GRUPOS DE RIESGO Y CÓMO SABER CUÁNDO AISLARLOS, ETC. MIS PRINCIPALES HERRAMIENTAS SON LA MINERÍA DE DATOS, LA ESTADÍSTICA Y LA OPTIMIZACIÓN COMBINATORIAL.



MAURICIO MONSALVE

Estudiante de Doctorado en Ciencia de la Computación, Universidad de Iowa, Estados Unidos. Ingeniero Civil en Computación y Magíster en Ciencias mención Computación, Universidad de Chile. Actualmente trabaja en el tema de epidemiología computacional y forma parte del grupo Computational Epidemiology (compEpi).

mauricio-monsalve@uiowa.edu

EPIDEMIOLOGÍA COMPUTACIONAL EN LA UNIVERSIDAD DE IOWA

La historia de cómo llegué a la Universidad de Iowa está ligada con Fulbright, y con la contingencia económica del momento. Postulé y quedé seleccionado en la beca de Doctorado Fulbright-Conicyt en el año 2008, mientras cursaba el Magíster en Ciencias de la Computación en el DCC de la Universidad de Chile. El mundo vivía una recesión económica generalizada en ese momento, por lo que las universidades norteamericanas estaban especialmente reticentes a entregar apoyo a estudiantes de otros Estados, que tienen que pagar más (el triple de arancel). Esto ponía una enorme carga al presupuesto que la beca entregaba en ese momento. Pero la Universidad de Iowa me daba una buena rebaja de arancel y una beca adicional. Acepté y me vine para Iowa City, Iowa, a mediados de 2009, justo después de defender la Tesis de Magíster.

Aprobando los cursos requeridos en el Programa de Doctorado, entré al grupo de Epidemiología Computacional, *compEpi*, de la Universidad. Éste es un grupo multidisciplinario, conformado principalmente por gente de Computación, pero también por gente de

Ingeniería Industrial, Bioestadística, Economía, Epidemiología y Salud Pública. Esta confluencia de gente permite un intercambio de ideas que siempre mantiene viva la motivación por hacer ciencia. Debido a que casi todos los edificios de la Universidad están cerca unos de otros, la formación de grupos interdisciplinarios es muy espontánea. Además, Iowa City es básicamente una ciudad-parque (urbanismo según el movimiento City Beautiful) con un grato ambiente cultural; es la tercera ciudad de la literatura según UNESCO.

Las actividades que realiza el grupo *compEpi* dependen principalmente de los intereses de sus participantes y del presupuesto del grupo. La gente de Epidemiología y Salud Pública pone los intereses generales en el grupo. La gente de Bioestadística está siempre interesada en probar nuevos métodos estadísticos en los datos disponibles. La gente de Economía está principalmente interesada en predecir las decisiones de los hospitales, combinando teoría de juegos y modelamiento epidémico. La gente de Ingeniería Industrial, está interesada en desarrollar aparatos no intrusivos con capacidades sensoriales. Y la gente de Computación es la más variada: unos trabajan con la Web y con comprensión textual (procesamiento de lenguaje natural); otros trabajan desarrollando aplicaciones para teléfonos celulares e inteligentes para monitorear síntomas, adherencia a políticas de vacunación, e incluso detectar enfermedades (hay un proyecto para clasificar el tipo de tos en las enfermedades); otros desarrollan métodos de monitoreo no intrusivos para detectar cuando un trabajador no adhiere al reglamento



hospitalario, como sería no lavarse las manos cuando debe, no lavarse bien las manos (hay técnicas para lavarse las manos), no usar guantes, etc. Y así hay muchos proyectos más.

Pese a que siempre hay muchos proyectos e intereses, la calidad y la cantidad de los proyectos está fuertemente limitada por el presupuesto. Claro, quisiéramos tener mucho dinero para tener los mejores datos y aparatos, y llevar a cabo cuánto experimento o desarrollo tengamos en mente. Pero el financiamiento es siempre limitado. Así que hay que salir a buscarlo. En efecto, muchas veces atendemos conferencias para lucirnos frente a gente que busca patrocinar proyectos. Si brillamos, nos conseguimos fondos para hacer más proyectos. Y temblamos cuando el presupuesto se empieza a acabar. ¡Si hasta los datos son caros! Los datos de mejor calidad (consistencia, limpieza, cantidad) valen varios miles de dólares. Por ello, la colección de datos por medio de la Web y los experimentos “en casa” se ven mucho más convenientes en general.

ESTUDIANDO LAS INFECCIONES INTRAHOSPITALARIAS

Pasando a mis temas de investigación, en el que he concentrado más esfuerzo ha sido en el tema de las HAIs o *healthcare-associated infections* (infecciones asociadas a la provisión de salud). En Estados Unidos solamente, alrededor de dos millones de personas adquieren HAIs más o menos serias cada año, muriendo varios miles. Me imagino que la situación en Chile no debe ser tan distinta, pero a escala. El problema de investigación es simple: entender cómo se contagian estas infecciones y qué se puede hacer para prevenirlas. Las HAIs actuales han evolucionado para ser resistentes a los antibióticos de consumo y a los externos (los antisépticos de limpieza). En la Universidad emplean un “robot” que rutinariamente esteriliza piezas desocupa-

das emitiendo fuerte radiación ultravioleta. Este enorme costo se ha vuelto necesario debido a lo resistentes y peligrosas que se han vuelto estas infecciones.

Uno de los proyectos en que rutinariamente incurre el grupo *compEpi* es el de usar sensores para monitorear el movimiento de los trabajadores de salud en alguna unidad del hospital de la Universidad, que es el hospital más importante de Iowa oriental. Estos despliegues caen en la categoría de *body area sensor networks* (redes de sensores en cuerpos) y *wireless sensor networks* (redes de sensores inalámbricas). He estado analizando los datos adquiridos en una medición hecha en la unidad de cuidados intensivos médicos del hospital, que principalmente recibe gente desde emergencias. El experimento duró diez días, incluyendo los turnos diurnos y nocturnos. Se monitoreó la ubicación de los trabajadores, así como la proximidad entre ellos y cuándo se lavaban las manos; ¡se colocaron sensores hasta en los dispensadores de jabón líquido!

Lo más interesante de este experimento son las posibilidades de investigación que salen de los datos. De partida, uno puede estudiar cuán efectiva es la higiene manual en la prevención de dispersión de infecciones. Usando los datos, uno puede simular que un paciente tiene una infección y que los trabajadores que entran en contacto con él pueden transmitirla a otros trabajadores y a otros pacientes. Luego, uno puede estudiar cuál es el nivel de higiene necesario (frecuencia del lavado de manos) para evitar el contagio de enfermedades y si se alcanzó. Pero es particularmente interesante saber que cambiar los niveles de higiene aporta muy poco cuando la higiene es ausente, tanto como cuando es prevalente.

Mi investigación ha sido principalmente con algo distinto, más social. He investigado los efectos que tiene la proximidad a colaboradores en el lavado de manos. Y he encontrado que, en efecto, los trabajadores son más propensos a lavarse las manos cuando están rodeados por otros trabajadores. Este es un tema interesante, muy relacionado a la teoría de juegos: es el efecto de los incentivos en el comportamiento de

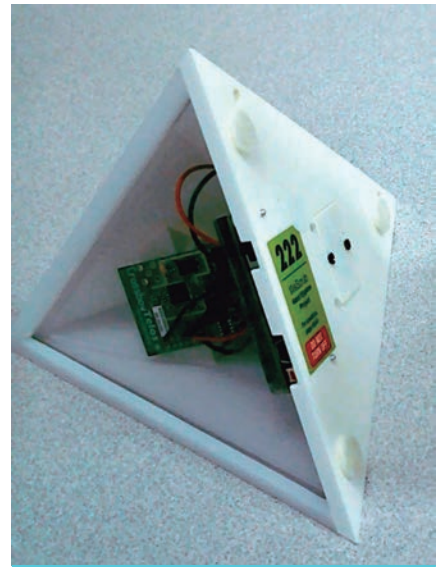


IMAGEN 1.

SENSOR FIJO UTILIZADO PARA UBICAR A LOS ESPECIALISTAS DE SALUD. LA FORMA DE PIRÁMIDE SE ELIGIÓ PARA NO IRRUMPIR CON LA DECORACIÓN EN LA UNIDAD.

las personas, medido experimentalmente. Los efectos son consistentes cuando controlo tanto por trabajador como por paciente. Lo bonito, analíticamente, es que esto lleva a una contradicción: rodearse por colaboradores mejora la higiene mientras que aumenta las probabilidades de transmitir las HAIs.

ENTRANDO EN COSAS ALGORÍTMICAS

También he investigado como extrapolar las redes de contacto de la Unidad de Cuidados Intensivos a todo el hospital. Esto se ve posible porque hay otros datos que ayudan a esto: los *logins* al sistema informático del hospital. Todos los trabajadores deben registrarse en este sistema cada vez que interactúan con un paciente. Esta regla rige desde oftalmología hasta cuidados intensivos postcirugía. Cada acceso a un computador es una muestra de dónde estuvo

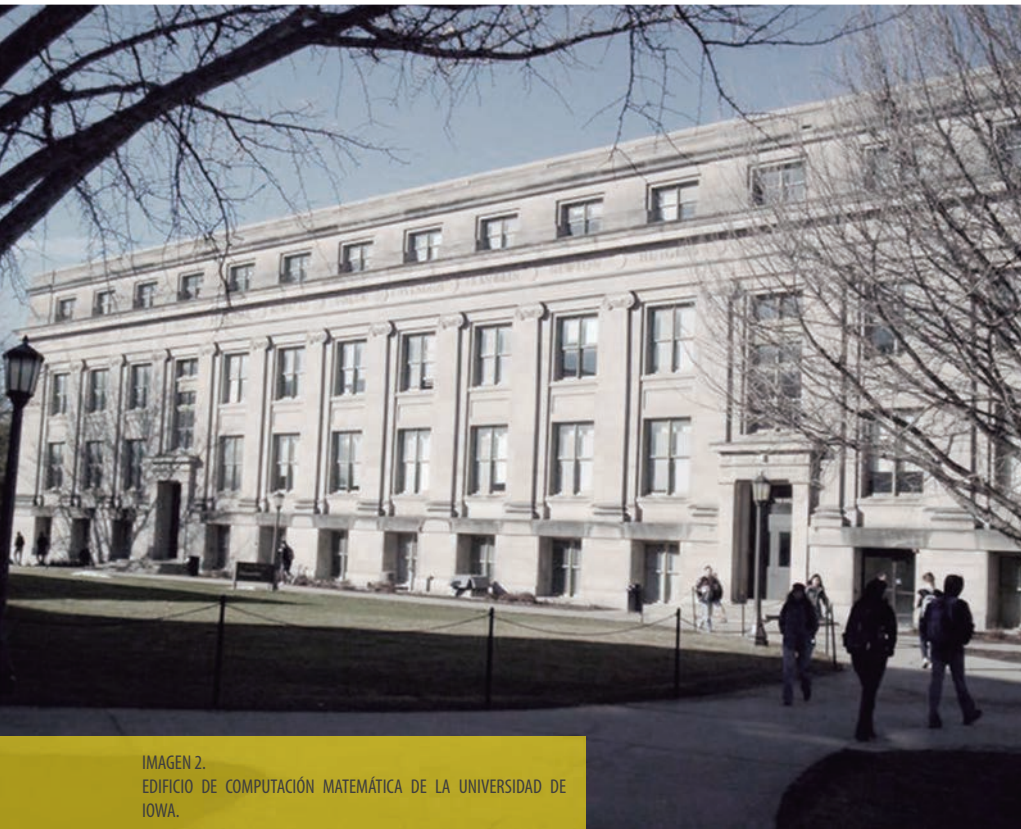


IMAGEN 2.
EDIFICIO DE COMPUTACIÓN MATEMÁTICA DE LA UNIVERSIDAD DE IOWA.

el trabajador en ese momento. Lo que falta es saber qué hace cuando no está en el computador. Bueno, conocer lo que ha ocurrido en la unidad de cuidados intensivos médicos es una pieza del rompecabezas. Juntando otras piezas más, sería posible estimar la topología de la red de contactos dentro el hospital por varios años usando la técnica de *predicción de vínculos*. Esta técnica está orientada a completar los grafos cuyos enlaces no han sido completamente recolectados. Y no interesa recuperar la red de contactos precisa de todo el hospital. Lo interesante es poder generar muestreos aleatorios de redes de contacto realistas que se parezcan a lo que realmente ocurrió en el hospital. Luego, es posible estudiar cómo las epidemias ocurren y qué hacer para prevenirlas, contenerlas y erradicarlas. Unos experimentos que hice mostraron que es posible re-estimar la red en la unidad de contactos intensivos a través de los accesos a computadores. Afortunadamente, hay algunos datos sobre el resto del hospital, y ahora el Grupo está llevando a cabo otra medición por sensores, esta vez en la unidad de diálisis.

¿Qué tan computacional ha sido esta investigación?

Bueno, para hacer los análisis ya mencionados, hice uso intensivo de grafos, minería de datos y simulación. Pero hay otra cosa que no he mencionado: que he tenido que procesar los datos de los sensores. Para maximizar la coherencia de los datos, he hecho uso de estructuras de datos que rápidamente generan “visitas” (cuando un trabajador está dentro de una pieza) para asignarlas a eventos de lavado de manos. De partida, los trabajadores se lavan las manos justo antes o justo después de atender a los pacientes. El objetivo es maximizar la cantidad de entradas o salidas de piezas que coinciden en tiempo, con cierta holgura, con el lavado de manos. Pero a su vez, cuando un dispensador de jabón se activa, muchos sensores registran ese evento. El trabajador más cercano sería el que se lava las manos, si no fuera porque muchas veces hay dos o tres trabajadores prácticamente equidistantes del dispensador. Entonces, también hay que desarrollar una regla para elegir ciega y consistentemente (ajustar los resultados caso



IMAGEN 3.

SENSOR INALÁMBRICO TELOS REV.B. ESTOS SENSORES FUERON USADOS PARA REGISTRAR LAS INTERACCIONES ENTRE ESPECIALISTAS DE SALUD CUANDO PRACTICABAN SANITIZACIÓN DE MANOS, DENTRO DE LA UNIDAD MÉDICA DE CUIDADOS INTENSIVOS DEL HOSPITAL DE LA UNIVERSIDAD DE IOWA.



IMAGEN 4.

SENSORES TELOS B EMBEBIDOS EN CARCASAS PLÁSTICAS DE LOCALIZADORES (PAGERS), PARA QUE LOS ESPECIALISTAS DE SALUD LOS PORTARAN.



IMAGEN 5. DISPENSADOR DE JABÓN-ALCOHOL PARA LA HIGIENE DE MANOS, EQUIPADO CON UN SENSOR PARA DETECTAR QUIÉN LO ACTIVO Y CUÁNDO LO HIZO.

a caso es fabricar resultados) qué trabajador se lavó las manos. Este problema de optimización ya cae en el mundo de la Inteligencia Artificial, y es difícil que alguien de otra especialidad lo resuelva.

MONITOREO DE ENFERMADES A GRAN ESCALA

Otra cosa que hice fue predecir la incidencia de la sífilis en Estados Unidos a través de la información recolectada por la Center for Disease Control and Prevention (Centro para el Control y Prevención de Enfermedades) y publicada en línea. El problema fue muy interesante, porque me encontré frente a datos de mala calidad, en cuanto eran incompletos (sesgados en los hospitales públicos) y no siempre veraces (errores de diagnóstico, que algunas veces se corregían).

Bajo la tesis de que Estados Unidos era un sistema complejo, empecé a ver las relaciones entre los Estados y la forma de las curvas epidémicas (prevalencia) en cada uno de ellos. Y desarrollé un método sencillo para predecir la incidencia en cada Estado a partir de información de otros Estados. Lo interesante es que este problema de series de tiempo no se había estudiado así en estadística. En efecto, fui en contra del método clásico de Box-Jenkins, que básicamente consiste en proyectar una serie de tiempo a través de diferencias finitas (si reemplazamos las series de tiempo por funciones analíticas unidimensionales, sería como encontrar relaciones lineales entre las derivadas de cualquier orden de cada función, para luego escribir un sistema de ecuaciones de diferencias que enlace todas las funciones. La proyección sería la curva resultante). Mi método era mucho menos analítico y mucho más computacional: literalmente, hice *matching* entre cada par de series de tiempo, viendo cuando las curvas se parecían más. Esto es, una correlación cruzada con retardos. Estoy pensando en mejorar este método por medio de abusar propiedades de transitividad: si X predice Y y Y predice Z , entonces puedo eliminar la

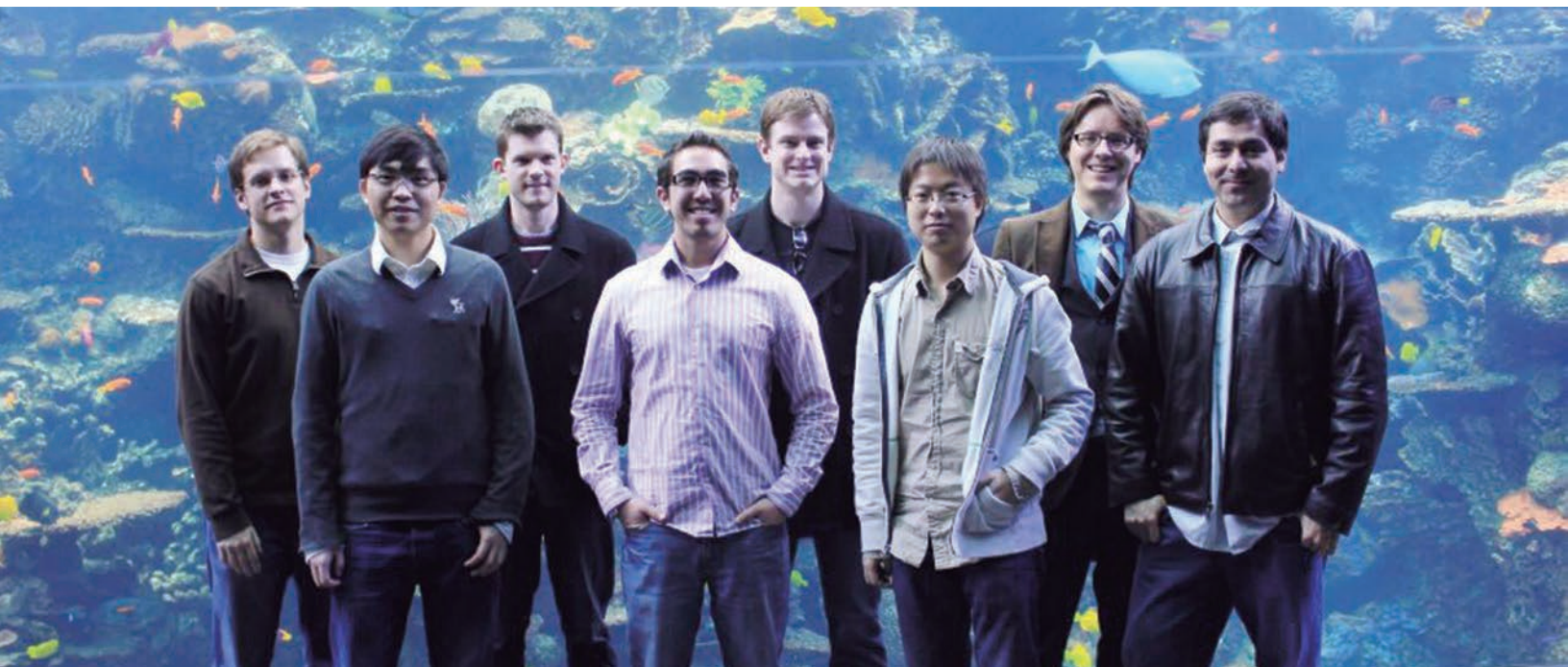


IMAGEN 6. GRUPO COMPEPI 2011 EN EL ACUARIO DE ATLANTA, GEORGIA. DE IZQUIERDA A DERECHA: ERIC FOSTER, ANSON TAI YATHO, AARON MILLER, LUCIO TOLENTINO, GEOFFREY FAIRCHILD, CHAO YANG, JASON FRIES Y MAURICIO MONSALVE. FOTO TOMADA DURANTE EL VIAJE A LA DÉCIMA CONFERENCIA DE LA INTERNATIONAL SOCIETY FOR DISEASE SURVEILLANCE (ISDS 2011).

relación X predice Z de mi tabla. Por supuesto, este problema es complicado porque tanto los retardos como las correlaciones son variados. Pero obtener una expresión mínima de predicción puede hacer más robusto el método en el tiempo y puede ilustrar qué ocurre realmente dentro del sistema complejo.

También quiero entrar a colaborar con otros en el grupo que se dedican a monitorear enfermedades a través de la Web. Los anuncios de sexo en las comunidades homosexuales tienden a estar correlacionados con los índices de enfermedades de transmisión sexual. Pero los anuncios además revelan los niveles de riesgo a los que están dispuestos quienes buscan sexo, pasando por no usar protección hasta usar drogas. Mezclando esto con las series de tiempo nacionales puede ser una buena manera de validar la calidad del monitoreo por la Web, así como usarlo como medio para conocer la sociología de las enfermedades de transmisión sexual. Pero la Web también permite el monitoreo de síntomas. Digamos, si una comunidad entera reporta jaquecas a través de Twitter, sabemos que algo está pasando. Google Trends también se puede usar para monitorear las búsquedas que tengan que ver con síntomas y consultas médicas. Entre más se descubra acerca del contagio de enfermedades, más se hace posible reconstruir qué está pasando; se pueden incluso construir redes de contacto usando información de la Web y descubrir cuánto de esa red está reflejada en la Web.

SALUD PÚBLICA COMPUTACIONAL

Para cerrar, quiero manifestar que es muy divertido trabajar con grupos multidisciplinarios. Las oportunidades para el descubrimiento científico se vuelven enormes. Pero así mismo, los desafíos crecen también. Hay que trabajar mucho más con los datos y los diseños experimentales para descubrir la información deseada. El método más barato, por ahora, es usar la Web, pero también es muy barato montar un computador con un micrófono o una cámara y registrar una red social. Lo segundo es mucho más costoso y a la gente no le gusta que la monitoreen.

Respecto a la salud pública, en Chile parece que hay muchos problemas que se pueden atacar usando métodos computacionales. Diagnósticos errados, colas de los hospitales (que no han acabado), mala repartición de recursos, burocracia, etcétera, son problemas que perfectamente pueden ser resueltos parcial o totalmente a través de buenos algoritmos y buenos sistemas de información. Esto es general para otros asuntos nacionales también. Sólo falta meter un buen poco de computación, curiosidad científica y voluntad pública para mitigarlos. ■



IMAGEN 7.
FOTO TOMADA DURANTE UNA PRESENTACIÓN EN LA CONFERENCIA
COMPLENET 2012, EN MELBOURNE, FLORIDA.