

Percolación de la epidemia de influenza AH1N1 en el mundo: Utilidad de los modelos predictivos basados en conectividad espacial

MAURICIO CANALS L.¹, ANDREA CANALS C.^a

Percolation of influenza AH1N1 epidemic in the world: Usefulness of the spatial-connectivity models

Background: The 2009 AH1N1 epidemics expanded rapidly around the world by the current connectivity conditions. The spread of epidemics can be described by the phenomenon of percolation, that allows the estimation of the threshold conditions that produce connectivity between different regions and that has been used to describe physical and ecological phenomena. **Aim:** To analyze the spread of AH1N1 epidemic based on information from the WHO. **Material and Methods:** The world was considered as composed of a set of countries and regular cells. The moment when the percolation occurred was analyzed and logistic regressions were adjusted to the change in the proportion of infected units versus time, comparing predicted and observed rates. **Results:** Percolation occurred in America on day 15, in Eurasia on day 32 and in the world on day 74. The models showed adequate predictive capacity. The predictions for the percolation of the epidemic in the world varied between days 66 and 75. The prediction based on countries was better than that based on cells. **Conclusions:** These results show that percolation theory fits well to the spread of epidemics. Predictions based only on data on-off (infected non infected) and in the progression of the proportion of infected cells are a good way of predicting the spread of an epidemic and when this crosses a region geographically.

(Rev Med Chile 2010; 138: 573-580).

Key words: Disease outbreaks; Influenza, human; World Health Organization.

¹Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile.

^aLicenciada en Matemáticas, Pontificia Universidad Católica de Chile.

Recibido el 4 de septiembre de 2009, aceptado el 31 de marzo de 2010.

Correspondencia a:
Mauricio Canals L.
Departamento de Ciencias Ecológicas, Facultad de Ciencias, Universidad de Chile
E-mail: mcanals@uchile.cl

La conectividad espacial de los procesos poblacionales es un tema actualmente muy relevante en ecología y en epidemiología¹. En el ámbito de la física, uno de los modelos que estudia la conectividad es la percolación^{2,3}; concepto fue introducido en 1956 por Hammersley y Broadbent para el estudio de la obstrucción de filtros de máscaras de gas³. La pregunta básica consistía en determinar el número o proporción de pequeños canales obstruidos que era necesario para determinar una obstrucción completa. Esta

pregunta aparentemente ajena a la medicina tiene gran relevancia en diversos campos donde interesa determinar el umbral necesario para conectividad entre unidades espaciales. Así, por ejemplo, el modelo de percolación ha sido utilizado para estudiar la continuidad espacial de la propagación de incendios, en la propagación de parásitos en árboles frutales y en la propagación de epidemias demostrando en este último campo gran utilidad, lo que inspiró a Grassberger a denominarlo la "geometría del contagio"^{2,4}. El modelo en su forma

más básica estudia la conectividad que se produce en una grilla de celdas cuadradas cuando cada celda se puede encontrar en dos estados ("off" y "on"). Así, es relevante la pregunta: ¿Qué proporción de celdas en estado "on" son necesarias para que se produzca una continuidad espacial de dichas celdas, permitiendo que el fenómeno atraviese de lado a lado la grilla?. En el caso que interesa en este artículo el estado "on" representa una unidad infectada y "off" lo contrario. Si suponemos que una unidad infectada puede contagiar a las vecinas que establecen continuidad con ella, se establece un sistema de propagación de la infección. Se puede demostrar que la percolación es un fenómeno de umbral y que para este caso particular el umbral de percolación es $p_u = 0,5927\dots$ ^{2,3,5}. Es decir bajo p_u la probabilidad de que el fenómeno cruce la grilla es muy baja, pero sobre p_u dicha probabilidad es muy alta. En la vecindad de p_u ($p - p_u \ll 1$) ocurren otros fenómenos muy interesantes como que los exponentes críticos β , γ y ν de las funciones características $P(p) = k_1 (p - p_c)^\beta$, $S(p) = k_2 |p - p_c|^{-\gamma}$, $L(p) = k_3 |p - p_c|^{-\nu}$ son universales (ie no dependen de la geometría de la grilla), donde $P(p)$ es la probabilidad que una celda pertenezca al grupo percolante, $S(p)$ es el tamaño promedio del grupo percolante y $L(p)$ es la probabilidad de aparición de un grupo percolante^{4,5}. El umbral de percolación es dependiente de la forma de las celdas de la grilla y de la existencia de puentes de continuidad entre celdas³.

Se ha propuesto que la percolación puede describir adecuadamente el avance geográfico de la epidemia AH1N1. Así, Canals (2009), basado en este modelo y utilizando los estados como celdas, propuso que la percolación describe adecuadamente el progreso que tuvo la epidemia en Estados Unidos de Norteamérica y que el umbral de percolación observado coincide con el esperado cuando se ajusta una regresión entre la proporción de estados infectados y el tiempo con el método de probito. También propuso, al inicio de la epidemia, basado en la utilización de países como celdas que la epidemia podría tener continuidad geográfica a los veintinueve días⁶. Aunque ni los estados de Estados Unidos ni los países del mundo constituyen una grilla cuadrada, los umbrales promedio de percolación en celdas de diversas figuras geométricas se asemejan a p_u de las grillas cuadradas⁶.

En este trabajo, proponemos que el progreso de la epidemia en el mundo se ajusta adecuadamente

al modelo de percolación basados en dos unidades geográficas: i) los países y b) celdas cuadradas entre paralelos y meridianos, y estudiamos la capacidad predictiva de los modelos.

Material y Métodos

La información fue obtenida de los reportes oficiales emitidos por la Organización Mundial de la Salud, durante el transcurso de la epidemia⁷ desde el 25 de abril al 12 de agosto de 2009. Se registraron en el curso de la epidemia los países con casos confirmados en función del tiempo. En este caso el país cambiaba de estado no-infectado a infectado. Se enmarcó además el mapa del mundo en una grilla de celdas cuadradas de 10 x 10 grados y se registró en el curso de la epidemia las celdas con casos confirmados. En el caso que un país registrara casos, todas las celdas que lo contenían cambiaban de estado no-infectado a infectado.

Las variables de interés fueron la proporción de países y de celdas infectadas en función del tiempo. Para calcular la proporción de países se consideraron los países y territorios según el estándar ISO3166-1 (n total = 269: 246 países + 23 territorios). Para calcular la proporción de celdas se consideraron las celdas que estaban ocupadas por continentes considerando el mínimo polígono convexo (n total = 368) (Figura 1).

Como definición operativa se consideró como percolación mundial al estado en el cual América y la unión Europa-Asia-Africa-Oceanía se tenían continuidad geográfica de celdas infectadas. Es decir, si se podía atravesar de extremo a extremo la grilla en los dos sentidos (Norte-Sur y Este-Oeste) en cada uno de estas dos unidades geográficas. No se consideró Groenlandia, ya que esta es una gran isla dependiente de Dinamarca pero con una ubicación geográfica más cercana a América del Norte.

Se realizaron regresiones logísticas entre estas proporciones con el tiempo y con el logaritmo del tiempo desde el inicio de la epidemia: Modelo lineal: $\log(P/(1-P)) = b_1 t + b_0$ y Modelo Log: $\log(P/(1-P)) = b_1 \log t + b_0$, donde P es la proporción de países (p) o de celdas (p_c), t el tiempo, b_1 la pendiente y b_0 el intercepto. Se calculó la transformada logito del umbral de percolación de las celdas ($p_{c_u} = 0,5927\dots$): $\log(0,5927/(1-0,5927)) = 0,1629$, y reemplazando este valor en las regresiones

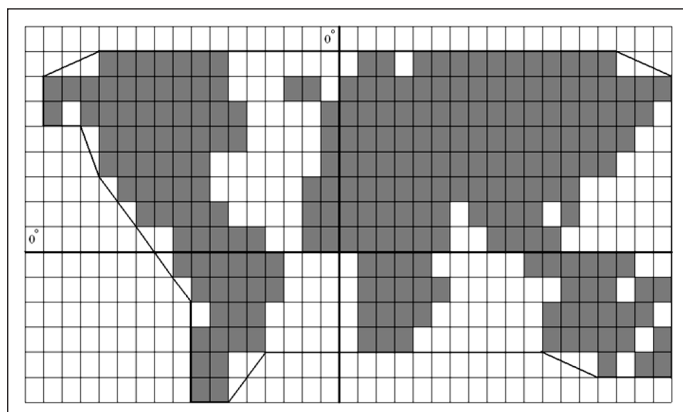


Figura 1. Grilla de celdas cuadradas enmarcando los continentes. Cada celda tiene 10° de latitud y 10° de longitud. Los ejes ortogonales indican el meridiano y el paralelo 0°. En gris las celdas que contienen tierras. Las líneas gruesas indican el polígono convexo considerado como "marco".

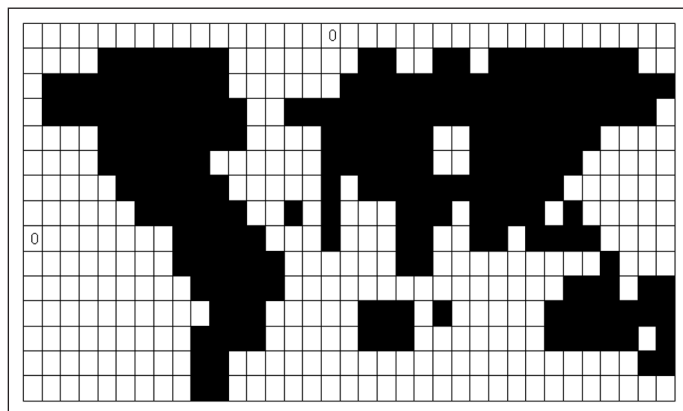


Figura 2. Celdas infectadas el día 7 de Julio de 2009. En este instante es posible establecer una continuidad geográfica (al nivel de celdas) atravesando de Norte a Sur y de Este a Oeste todo el mundo.

se determinó el tiempo predicho para alcanzar dicho umbral (t_p). Cuando usamos el modelo logarítmico se obtiene una predicción $Lt_p = \log(t_p)$ y entonces la predicción es $t_p = \text{antilog}(Lt_p)$. Para esta predicción se estimó el intervalo del 95% confianza mediante: $IC_{0,95} = (\log(t_p) \pm t_{(1-\alpha/2)} Es(\log t_p))$ y luego se calcularon los antilog de los extremos. Ya que el umbral de percolación de celdas infectadas no equivale a la misma proporción de países infectados, se estableció la relación entre estas proporciones con análisis de regresión al origen: $pc = b_1 p$, donde pc es la proporción de celdas y p la de países. Así se obtuvo el umbral de percolación para la proporción de países infectados: $p_u = pc_u / b_1$ y con la misma lógica anterior se determinó el tiempo predicho de percolación (ahora basado en los países). Este procedimiento se realizó usando la

información completa (110 días) y considerando los 10, 15, 20, 25 y 30 primeros días de la epidemia con el fin de estudiar la capacidad predictiva de los modelos.

Resultados

La percolación en América ocurrió en el día 15; en Eurasia en el día 32 y en el mundo en el día 7 de julio que equivale al día 74 desde el inicio de la epidemia (Figura 2). En este día había cambiado de estado 58,97% de las celdas, lo que se ajusta muy bien al valor esperado (59,27%). La proporción de países infectados en esta fecha fue menor, aunque también cercano al umbral de percolación: 51,30%. La percolación definida sobre la base de las celdas fue adecuadamente representativa

de la situación ocurrida en los países (Figura 3). Tanto el Modelo lineal como el Modelo Log para la proporción de países y para la proporción de celdas tuvieron un ajuste adecuado, sin embargo el ajuste fue mucho mejor utilizando el modelo Log (Tabla 1). Los modelos lineales presentaron una distorsión importante en los primeros días, donde los valores observados cayeron por debajo de los valores predichos lo que reduce su capacidad predictiva (Figuras 4 y 5).

La regresión entre p_c y p fue: $p_c = 1,538 p$ ($R^2 = 0,829$, $F_{1,43} = 215,0$, $p \ll 0,001$). Así, la proporción umbral de países que debería determinar percolación fue: $p_u = 0,5927/1,538 = 0,385$, y su

transformación logito fue $\log(0,385/(1-0,385)) = -0,203$.

Los valores predichos para la percolación calculados con el modelo Log fueron muy buenos tanto utilizando p como utilizando p_c (Tabla 2). Al utilizar los 110 días de información los valores predichos son de 66,2 y 75,54 días utilizando p y p_c respectivamente, lo que se ajusta muy bien al valor observado (74 días). La capacidad predictiva del modelo Log utilizando la proporción de celdas (p_c) solo logra una adecuada capacidad predictiva con 25 días de información, mientras que la capacidad predictiva de la proporción de países es excelente y estable desde el inicio.



Figura 3. Países infectados el día 7 de Julio de 2009.

Tabla 1. Regresiones logísticas entre la proporción de países infectados y el tiempo y entre la proporción de celdas y el tiempo para el progreso de la epidemia AH1N1 en el mundo

Tipo de modelo	Modelo	R2	F1,42
Lineal-países	$\log\left(\frac{p}{1-p}\right) = 0,018 \cdot t - 1,35$	0,79	163,0*
Lineal-celdas	$\log\left(\frac{p_c}{1-p_c}\right) = 0,047 \cdot t - 0,61$	0,66	84,6*
Log-países	$\log\left(\frac{p}{1-p}\right) = 1,168 \cdot \log t - 2,33$	0,97	1.387,5*
Log-celdas	$\log\left(\frac{p_c}{1-p_c}\right) = 0,773 \cdot \log t - 1,29$	0,94	656,4*

El símbolo * indica $p < 0,01$.

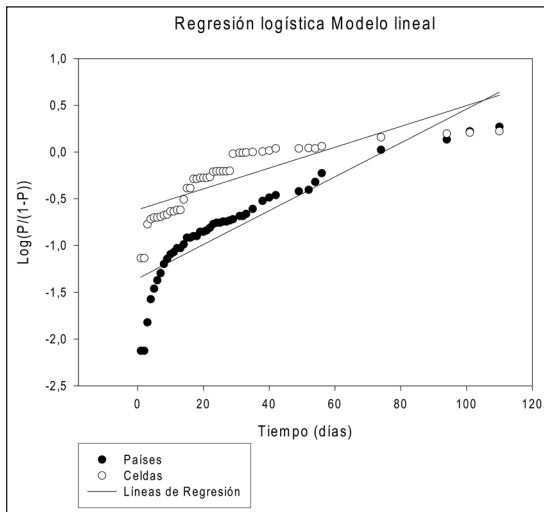


Figura 4. Regresión logística entre las proporciones de países y celdas en función del tiempo. La ordenada corresponde a la transformación logito ($\log(P/(1-P))$), donde P es la proporción de países (p) o la proporción de celdas (pc).

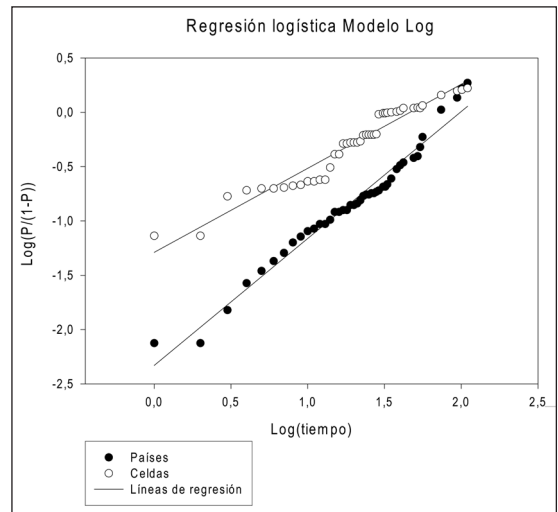


Figura 5. Regresión logística entre las proporciones de países y celdas en función del logaritmo del tiempo. La ordenada corresponde a la transformación logito ($\log(P/(1-P))$), donde P es la proporción de países (p) o la proporción de celdas (pc).

Tabla 2. Capacidad predictiva del tiempo de percolación (t_p) y su intervalo del 95% de confianza ($IC_{0,95} t_p$) para las regresiones logísticas entre la proporción de países infectados y el logaritmo del tiempo (Modelo Log) y entre la proporción de celdas y el logaritmo del tiempo considerando diferentes números de días desde el inicio de la epidemia

Países	Pendiente	Intercepto	R2	F	t_p	$IC_{0,95} t_p$
N = 10 días	1,173	-2,29	0,94	136,9*	63,23	38,6 - 103,5
N = 15 días	1,152	-2,28	0,96	366,9*	63,56	33,5 - 120,6
N = 20 día	1,116	-2,22	0,97	632,8*	69,73	33,3 - 145,9
N = 25 días	1,090	-2,24	0,98	947,7*	74,43	34,0 - 162,7
N = 30 días	1,060	-2,22	0,98	1197,4*	80,09	34,6 - 185,3
N = 110 días	1,168	-2,33	0,97	1387,5*	66,18	46,9 - 94,6
Celdas						
N = 10 días	0,542	-1,14	0,81	39,4*	253,95	79,8 - 807,8
N = 15 días	0,534	-1,14	0,84	72,5*	274,34	101,4 - 742,3
N = 20 días	0,652	-1,21	0,86	115,4*	125,82	63,1 - 250,9
N = 25 días	0,700	-1,23	0,89	197,1*	100,14	54,5 - 184,1
N = 30 días	0,746	-1,27	0,90	257,5*	83,76	52,7 - 166,8
N = 110 días	0,773	-1,29	0,94	656,4*	75,54	45,8 - 124,6

El símbolo * indica $p < 0,01$.

Discusión

Uno de los aspectos transcendentales en epidemiología es intentar la predicción de la evolución de las enfermedades infecciosas. Este intento se realiza habitualmente mediante modelos matemáticos que consideran el progreso de casos en el tiempo en una cierta localidad y parámetros relevantes como el número reproductivo, tiempo de doblamiento, transmisibilidad, etc⁸⁻¹⁷. Sin embargo, estos modelos no incluyen la dimensión espacial del progreso de la enfermedad ni las condiciones de conectividad que se establecen entre regiones distantes, factores que actualmente son muy relevantes en la ecología, epidemiología y en otros ámbitos como la física^{2-4,18}. La percolación, siendo un fenómeno de umbral, permite explicar cambios bruscos en la propagación, sea esta de agua, gases, parásitos o epidemias^{2,4} y es por tanto relevante en la propagación de la actual epidemia AH1N1. Mientras que se ha reportado que epidemias previas tomaron muchos meses y años en adquirir el carácter de pandemias¹⁹, esta lo ha hecho rápidamente en menos de tres meses²⁰.

El modelo de percolación muestra un muy buen ajuste al progreso de la actual epidemia AH1N1 en el mundo. En trabajos previos se ha reportado una situación similar, a nivel local, para el progreso de la epidemia en Estados Unidos de Norteamérica de Norteamérica. También usando un modelo basado en la transformada de probito y a sólo diecisiete días del inicio de la epidemia se estimó una posible percolación mundial a los 29 días⁶, lo que constituye una subestimación de lo que finalmente sucedió. Sin embargo, en estos trabajos se advierte que los umbrales de percolación varían dependiendo de las formas geométricas de las celdas y es claro que las formas irregulares y variables de los países distan bastante de celdas cuadradas. La evolución de la epidemia en Estados Unidos sin embargo tenía un muy buen ajuste explicado por la particular disposición de los estados que prácticamente constituyen una red continua⁶.

Considerando las celdas como unidades geográficas obtuvimos una coincidencia prácticamente exacta entre el umbral de percolación esperado ($p_u = 0,5927\dots$) y lo observado ($p_{c_u} = 0,5897$), lo que simplemente comprueba empíricamente lo que matemáticamente se ha calculado para grillas cuadradas bidimensionales⁵. La única diferencia que reviste este caso es que el contorno de la grilla

es un polígono y que dentro de la grilla hay zonas que no pueden ser ocupadas, ya que en ellas no hay tierra. La utilización de celdas regulariza la forma y tamaño de las unidades geográficas y establece continuidad donde en realidad no la hay. Por ejemplo entre Indonesia, Papua-Nueva Guinea, Australia y la Micronesia no hay una continuidad real, pero ésta se establece cuando cambiamos de estado toda la celda donde se encuentra el país infectado. En este sentido es una representación en "grano grueso" de la situación, pero que sin embargo, tiene la virtud que al regularizar la forma y tamaño de las unidades permite la aplicabilidad de la teoría, lo que explica el excelente ajuste. Sin embargo, no es lo mismo proporción de celdas que proporción de países y entonces cabe preguntarse si el estado de percolación de celdas es equivalente a la continuidad geográfica entre los países. Las Figuras 2 y 3 comparan la situación entre celdas y países en el día en que ocurrió la percolación. En ellas se puede ver que efectivamente hay una continuidad en toda América y en la unidad Eurasia-Africa-Oceanía. En esta última es posible atravesar de Este a Oeste por Eurasia y de Norte a Sur a través del Sudeste de Asia-Oceanía, aunque no así por Africa. Aquí es importante notar que este no es un requisito para la percolación, ya que esta se produce por la continuidad entre dos puntos cualesquiera de los límites opuestos de la grilla.

El ajuste de las regresiones logísticas fue muy bueno, especialmente cuando usamos el logaritmo del tiempo como variable independiente. Cuando se usa el tiempo directamente, aunque tiene un ajuste relativamente adecuado, se produce una distorsión en los primeros días cuando los valores observados caen por debajo de la recta ajustada lo que reduce drásticamente su capacidad predictiva. Este resultado explica la subestimación del tiempo de percolación de trabajos previos⁶.

Cuando usamos el modelo Log tanto para la proporción de países como para celdas la predicción del tiempo de percolación fue muy bueno (66,2 y 75,5 días vs 74 días), sin embargo, esto sólo indica un buen ajuste cuando se utiliza toda la información. Es decir cuando ya ha ocurrido el fenómeno. Cuando comparamos los modelos (países y celdas) y la capacidad predictiva de ellos a los 10, 15, 20, 25 y 30 días del inicio de la epidemia se observa que el ajuste es mejor al usar los países como unidades. En este caso el ascenso en la proporción de países es relativamente

continuo, mientras que para el caso de las celdas, este se produce “en escalera”. Este efecto se debe a que en muchas ocasiones, aunque aumente el número de países infectados el número de celdas puede no variar ya que los países que declaran nuevos casos caen en celdas ya ocupadas. Así, no es lo mismo “p” que “pc” y son diferentes los umbrales de percolación. La relación entre p y pc estimada a través de regresión permitió estimar la proporción umbral de países para desencadenar la percolación. Esta fue 0,385, es decir mientras que se necesitan 59,27% de celdas para desencadenar la percolación, sólo bastaría 38,5% de los países. La diferente evolución de p y pc en el tiempo y las diferencias de ajuste se reflejaron en la capacidad predictiva de los modelos. Mientras que usando las celdas como unidades se obtiene una adecuada capacidad predictiva recién a los 30 días de iniciada la epidemia (Tabla 2), cuando usamos los países, después de calcular el umbral ($p_u = 0,385$), la predicción es excelente desde los primeros días y además muestra una gran estabilidad, permaneciendo las predicciones y los parámetros del modelo prácticamente constante. Los intervalos de confianza van reduciendo su amplitud en la medida que se tiene mayor información (más días), llegando a una precisión máxima de 1,6 a 2 meses según el modelo. Así mientras las celdas permiten una mejor definición de la percolación a nivel global, los países son unidades de “grano más fino” que permite una mejor predictibilidad del fenómeno después de ajustar la relación entre proporción de países y celdas.

A diferencia de los modelos epidemiológicos clásicos que usan ecuaciones diferenciales o ecuaciones de diferencia (ie modelos SEIR)⁹ para estimar el número de casos en un tiempo determinado, el modelo aquí mostrado está basado en la propagación espacial incorporando al tiempo como una variable que permita estimar cuando ocurrirá la percolación. Es decir aquél instante en que desaparecen las barreras geográficas que podrían evitar la propagación a todo el mundo. Los modelos SEIR (Susceptibles, Enfermos, Infectados, Recuperados) permiten estimar rápidamente el número de casos, especialmente a corto plazo en una unidad geográfica determinada, pero no informan de la propagación hacia otras unidades vecinas. En cambio el modelo presentado permite predecir la propagación, pero no informa de la cantidad de casos en cada unidad geográfica

afectada. En este sentido, ambos enfoques son complementarios y ambas informaciones son relevantes a la hora de decidir medidas de contención de una epidemia.

El adecuado ajuste y buena capacidad predictiva de nuestro modelo permiten proponerlo como un método epidemiológico para seguir y predecir el avance geográfico de enfermedades infecciosas adaptando el tamaño de las celdas a la situación local, por ejemplo comunas y ciudades.

Referencias

1. Marquet PA. Metapopulations. En *Encyclopedia of global environmental change. The earth system: biological and ecological dimensions of global environmental change*. Munn T Editor. Chichester, United States: John Wiley & Sons, Chichester; 2002: 411-20.
2. Grassberger P. On the critical behavior of the general epidemic process and dimensional percolation. *Mathematical Bioscience* 1983; 63 : 157-72.
3. Grassberger P. La percolation ou la geometrie de la contagion. *La Recherche* 1991; 22 : 640-6.
4. Miller JC. Percolation and epidemics in clustered networks. *Physical Reviews* 2009; E80: 020901-5.
5. Feng X, Deng Y, Blote HWJ. Percolation transition in two dimensions. *Physical Review* 2008; E78: 031136.
6. Canals M. Inicio de la pandemia AH1N1: Algebra, cálculo y geometría del contagio. *Rev Med Chile* 2009; 137: 852-6.
7. World Health Organization. 2009. <http://www.who.int/en/>
8. Canals M, Cattán PE. *Zoología médica: una visión de las especies potencialmente peligrosas desde la perspectiva de la biodiversidad. Generalidades y protozoos*. Santiago, Chile: Editorial Universitaria; 2006, p. 160.
9. Canals M. Predictibilidad a corto plazo de casos de influenza AH1N1 basado en modelos determinísticos. *Rev Chil Infectol* 2010; 27: 114-20.
10. Anderson RM, May RM. Population biology of infectious diseases: Part I. *Nature* 1979; 280 (5721): 361-7.
11. Hefferman JM, Smith RJ, Wahl LM. Perspectives on the basic reproductive ratio. *J R Soc Interface* 2005; doi: 10.1098/rsif.2005.0042: 1-13.
12. Bailey N. *The mathematical theory of infectious diseases and its applications*. London, England:Griffin; 1975. p. 413.
13. Canals M, Martínez L, Firinguetti L. Dinámica ecológica de enfermedades infecciosas. I parte: Estacionalidad. *Rev Med Chile* 1989; 117: 87-94.

14. Canals M, Martínez L, Firinguetti L. Dinámica ecológica de enfermedades infecciosas. II parte: Recurrencia epidémica. Rev Med Chile 1989; 117: 95-100.
15. Canals M. Sarampión en Chile: Una peligrosa onda. Rev Med Chile 1992; 120: 585-8.
16. Canals M. Dinámica de transmisión de enfermedades infecciosas: una expresión particular de una dinámica depredador-presa. Rev Med Chile 1993; 121: 1316-20.
17. May R, Anderson RM. Population biology of infectious diseases: Part II. Nature 1979; 280: 455-61.
18. Solé R. En *Redes complejas. Del genoma a internet*. Solé R Editor. Barcelona, España: Tusquets; 2009. p. 86.
19. Acha PN, Cifres B. Zoonosis y enfermedades transmisibles comunes al hombre y los animales. II Clamidiosis, rickettsiosis y virosis. OPS publicación científica y técnica 2003; 580: 329-46.
20. Fraser C, Donnelly CA, Cauchemez S, Hanage WP, Van Kerkhove MD, Hollingsworth TD, et al. Pandemic Potential of a Strain of Influenza A (H1N1): Early Finding Science 2009; 324 (5934): 1557-61.